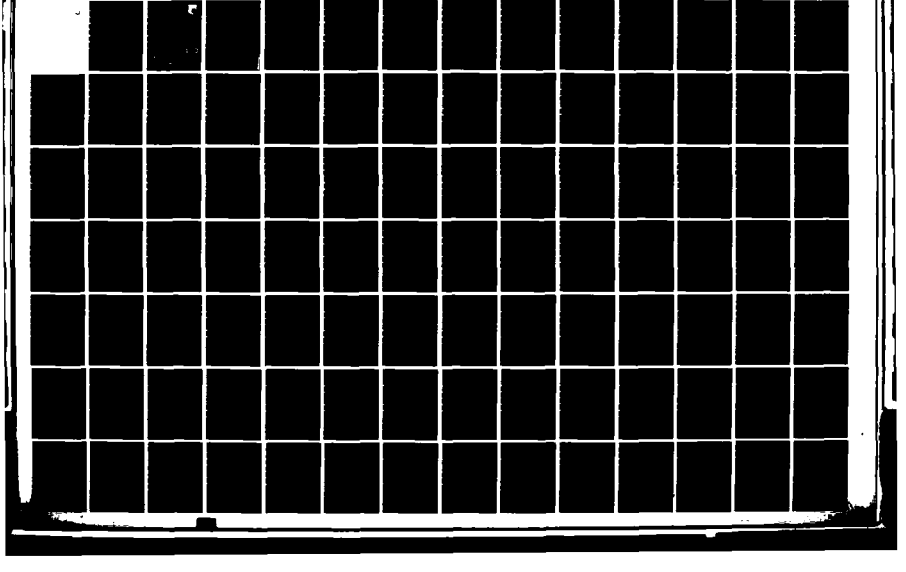
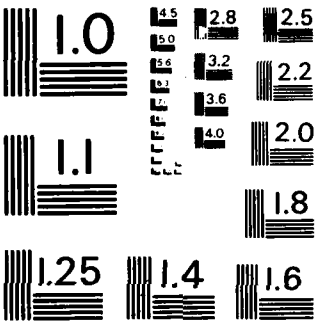
AD-A142 001 FLAW DETECTION RELIABILITY CRITERIA VOLUME 1 METHODS 1/2
AND RESULTS(U) DAYTON UNIV OH RESEARCH INST
A P BERENS ET AL. APR 84 UDR-TR-83-137-VOL-1
UNCLASSIFIED AFWAL-TR-84-4022-VOL-1 F33615-82-C-5030 F/G 14/2 NL





AFWAL-TR-84-4022
Volume I

FLAW DETECTION RELIABILITY CRITERIA

Volume I - Methods and Results

A. P. Berens
P. W. Hovey

University of Dayton
Research Institute

April 1984
Final Report for Period May 1982 - September 1983

MATERIALS LABORATORY
AIR FORCE WRIGHT-AERONAUTICAL LABORATORIES
AIR FORCE SYSTEM COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture use, or sell any patented invention that may in any way be related thereto.

This report has been reviewed by the Office of Public Affairs (ASD/PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

Joseph A. Moyzis

JOSEPH A. MOYZIS, JR.
Technical Area Manager
Nondestructive Evaluation Branch
Metals and Ceramics Division
FOR THE COMMANDER

D. M. Forney

D. M. FORNEY, JR., Chief
Nondestructive Evaluation Branch
Metals and Ceramics Division

"If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify AFWAL/MLLP, W-PAFB, OH 45433 to help us maintain a current mailing list".

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFWAL-TR-84-4022, VOLUME I	2. GOVT ACCESSION NO.: AD-A152001
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE *(and Subtitle)*: FLAW DETECTION RELIABILITY CRITERIA Volume I - Methods and Results	5. TYPE OF REPORT & PERIOD COVERED: Final May 1982 - September 1983
	6. PERFORMING ORG. REPORT NUMBER: UDR-TR-83-137
7. AUTHOR(s): A. P. Berens and P. W. Hovey	8. CONTRACT OR GRANT NUMBER(s): F33615-82-C-5030
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of Dayton Research Institute Dayton, Ohio 45469	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: P. E. 62102F 2418-05-31
11. CONTROLLING OFFICE NAME AND ADDRESS: Materials Laboratory (AFWAL/MLLP) Air Force Wright Aeronautical Laboratories Wright-Patterson AFB, OH 45433	12. REPORT DATE: April 1984 13. NUMBER OF PAGES: 166
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*:	15. SECURITY CLASS. *(of this report)*: Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

AFWAL-TR-84-4022, Volume II contains computer software, therefore distribution is limited in accordance with AFR 300-6 (DOD Dir. 4160.19, dated 5 April 1973). Non-DOD requests must include the statement of terms and conditions contained in Attachment 21 of AFR 300-6.

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

non-destructive inspection	crack detection reliability
NDI reliability	damage tolerance analysis
Inspection capability	initial flaw sizes
Probability of crack detection	inspection limits

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Air Force inspection limits for damage tolerance analyses have been defined as those crack lengths for which there is 95 percent confidence that at least 90 percent of all cracks will be detected. This definition of inspection limit cannot be used in structural risk assessments as it is not linked to the probability of missing the flaws that may be present in a structure. This program investigated alternate methods of defining flaw detection reliability criteria and, in particular, focuses on a characterization based on a fixed probability of having cracks greater than the inspection limit after the inspection.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

A major part of the program was devoted to methods of estimating the POD function from inspection reliability demonstration experiments. These methods are extensively discussed.

FOREWORD

This technical report summarizes the work performed by the University of Dayton Research Institute under Materials Laboratory Contract Number F33615-82-C-5030. The work was performed between May 1982 and September 1983. Dr. Joseph A. Moyzis of the Materials Laboratory was the Air Force Project Monitor and Dr. Alan P. Berens of the University of Dayton was the Principal Investigator.

The final report of this work comprises two volumes. A complete description of the methods and results of the program are contained in Volume I. As part of the study, two computer programs were written: a) a program for calculating effective inspection reliability and exceedance probabilities for selected POD functions and crack length distributions, and b) a program for simulating NDI experiments and estimating the parameters of the POD function. A user's manual and the listings for these programs are contained in Volume II.

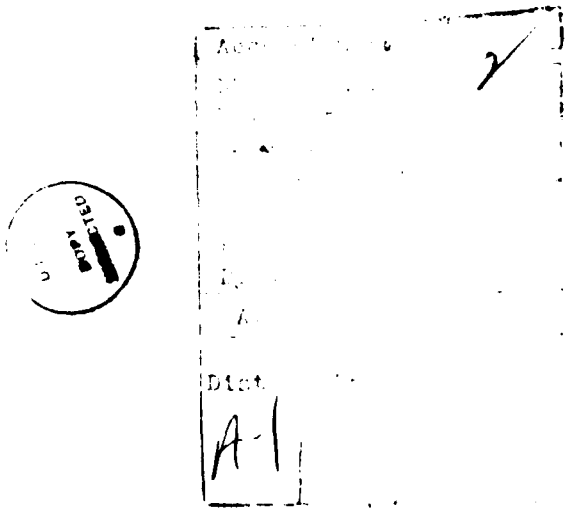


TABLE OF CONTENTS

TABLE OF CONTENTS (Concluded)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

LIST OF ILLUSTRATIONS (Continued)

LIST OF ILLUSTRATIONS (Continued)

LIST OF ILLUSTRATIONS (Continued)

LIST OF ILLUSTRATIONS (Concluded)

LIST OF TABLES

SECTION 1
INTRODUCTION

Guaranteeing structural safety through damage tolerance analysis has focused attention on quantifying the capability of nondestructive inspection (NDI) systems. To predict the growth of the largest flaw that might be present at a critical location requires knowledge of the flaw sizes that might be missed at an inspection. However, factors beyond the control of the NDI system influence the detectability of flaws at the small sizes of interest and NDI capability must be characterized in probabilistic terms. For a particular application of a given NDI system, capability is defined as the probability of crack detection (POD) of all cracks of fixed length, i.e., in terms of the function POD(a).

A one number characterization of NDI capability is currently in use as defined by the minimum crack length for which there is a fixed degree of confidence (CL) that at least a fixed proportion (POD) of cracks will be detected, i.e., the POD/CL crack length. For example, there is 95 percent confidence that at least 90 percent of all cracks greater than the 90/95 crack length will be detected. Several studies in recent years have resulted in a different approach to characterizing NDI capability and have demonstrated weaknesses in the POD/CL type characterization.

In the late 1970's, results from multiple, independent inspections of the same fatigue cracks were generated under an Air Force NDI reliability program known as the "Have Cracks Will Travel" Program[1]. These data demonstrated conclusively that different cracks of the same length can have significantly different crack detection probabilities. This realization led to a regression model approach for estimating the POD(a) function and, further, permitted realistic computer simulations of NDI experiments. The simulations, in turn, demonstrated the extreme statistical instability of the POD/CL type characterization [2]. This instability implies that if an NDI reliability demonstration program were repeated under rigidly controlled conditions, the 90/95 crack lengths could be radically different due only to the statistical variability of inspection results.

Further, the relation between a 90/9 crack length and the chances of not detecting a longer crack is entirely unknown.

For damage tolerance calculations at the design stages of aircraft development, the Air Force has circumvented the probabilistic aspects of initial crack sizes by specifying the crack lengths which must be used for new and in-service inspected structures. These numbers are only vaguely related to current NDI capability as they were established on the basis of a broad brush evaluation of pre-1974 inspection capability and are also considered to cover manufacturing or repair defects which are not necessarily detectable by a specific NDI system. However, the specification does contain a provision for lowering the initial flaw size assumptions for one category of virgin structure based on a demonstration that the NDI system to be used has a lower 90/95 crack length.

In view of the apparent difficulties with the 90/95 crack length type of NDI reliability characterization, this program was undertaken with two objectives: a) to determine a procedure for specifying crack size assumptions in damage tolerance analyses which correlates NDI capability with the risk of structural failure; and, b) to review existing Air Force crack size assumptions and make recommendations for incorporating the new procedure. To achieve these objectives considerable effort was devoted to developing and evaluating methods for analyzing NDI reliability data.

This report comprises 6 Sections. In Section 2, the basis for correlating inspection uncertainty with risk of structural failure is established. This section emphasizes the need to account for both the sizes of the cracks in the population of structural details and the probability of crack detection as a function of crack size. Section 3 presents an extensive discussion and mathematical details related to the analysis of NDI reliability data. Section 4 presents a sensitivity analysis of reliability characterizations and an evaluation of the effect of the sampling variation in the estimates of the POD function parameters on the probability of missing large cracks at an inspection. This section

ties the material of Sections 2 and 3. Section 5 discusses the Air Force Aircraft Structural Integrity Program (ASIP) ramifications of NDI reliability characterizations and presents recommendations for using the methods of this report in determining initial flaw size assumptions in damage tolerance analyses. Conclusions are presented in Section 6.

SECTION 2
ANALYTICAL FRAMEWORK

The damage tolerance approach to structural safety is centered on a philosophy of insuring safe operation in the presence of flaws. Accordingly, flaws are assumed to be present at all critical locations and it is demonstrated that these flaws will not grow to a critical size in the usage environment during each period of operation. This process requires bounds on the sizes of the flaws that may be in the structure and the bounds must be quantified in terms amenable to analysis. Since fracture mechanics methods are used to predict flaw growth, flaw sizes are stated in terms of a crack length or an equivalent crack length. Manufacturing quality control and field inspections are intended to detect all flaws but the capability of current inspection systems can only be expressed in probabilistic terms at the small crack sizes of interest. Since no guarantee can be given that all flaws greater than a predetermined bound will be detected, the inspection element contributes to the overall failure risk of the damage tolerance process.

This section presents a simple approach to measuring the effect of inspection uncertainty on structural integrity by correlating the NDI system with the distribution of crack sizes in a structure after inspection. This approach permits determining that crack size for which there is a predetermined probability of exceedance. Since the calculations require knowledge of the POD function and the distribution of crack sizes in the structure, these topics are also discussed.

2.1 QUANTIFYING NDI UNCERTAINTY FOR DAMAGE TOLERANCE ANALYSIS

The Air Force application of damage tolerance analysis is usually summarized as shown in the schematic of Figure 1. The as-manufactured or virgin structural detail is assumed to have a flaw of length a_o. The growth of this flaw is predicted for design usage and will reach critical crack size, a_f, after t_D flight hours. Since the Air Force philosophy calls for an inspection at half the time required for the potential crack to reach critical, the first

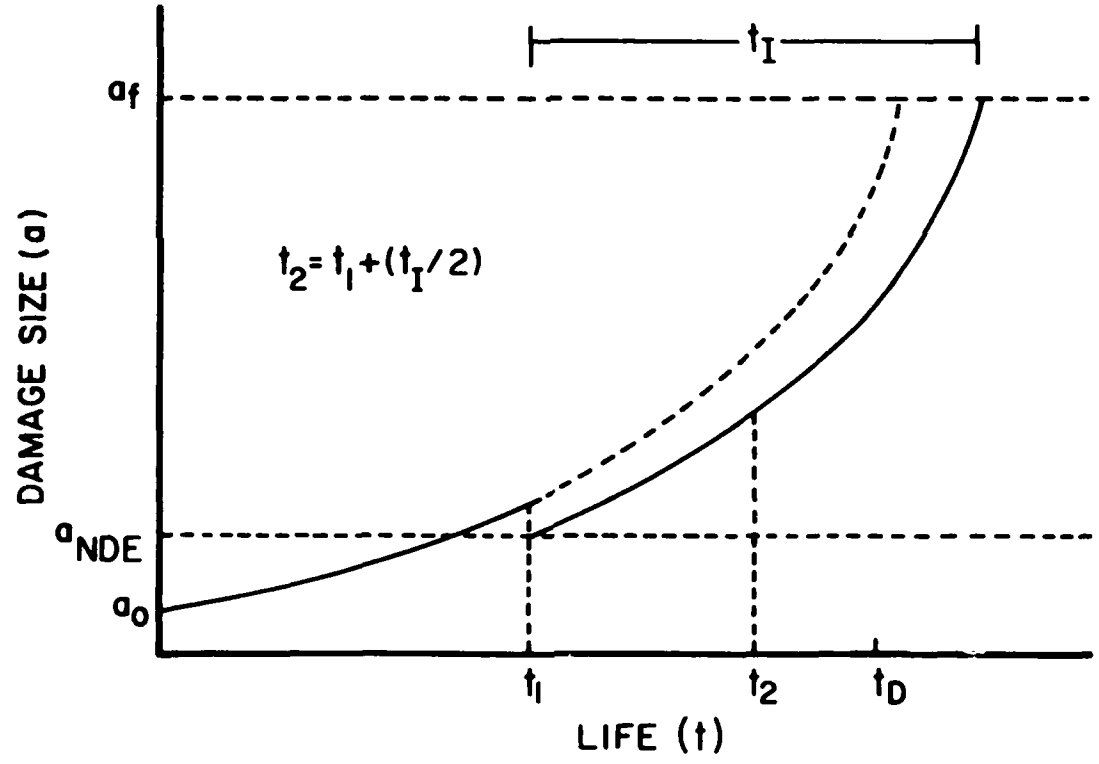


Figure 1. Crack Growth-Life Curve to Second Inspection.

inspection is scheduled at $t_1 = t_D/2$ flight hours. The maintenance action is assumed to find and remove any flaw of length greater than a_{NDE} so that a_{NDE} is the initial crack size for the next usage period. Since a crack of length a_{NDE} is predicted to reach critical size in t_I flight hours, the second inspection is scheduled at $t_2 = t_1 + t_I/2$. The process continues until the structure becomes uneconomical to maintain.

The application of this process requires, a) extensive knowledge concerning the location, geometry, and material properties of all potentially critical areas of the structure; b) predicted stress histories for each critical location during the service usage period; and c) bounds on the sizes of potential flaws. The first two of these factors are highly aircraft dependent as they can be determined only for a specific design and usage. The third factor is generally considered to be a measure of inspection capability and/or an equivalent flaw size which covers possible defects resulting from the manufacture or maintenance process [3].

Ostensibly, damage tolerance analysis is deterministic since all potential flaw initiation and growth is assumed to be bounded by the crack growth-life curve. In actuality, there are several non-deterministic or random elements in the process which result in a positive (albeit very small) probability of failure. These include the random nature of the actual stresses to be encountered, inspection uncertainty and the variability inherent in material properties. At present, there are insufficient data to combine all of the random elements and the potential joint effects of adjacent structural details in a generally accepted model to realistically predict the structural failure probabilities. The Air Force accounts for the uncertainty by a) tracking the severity of usage of each individual aircraft in an attempt to account for variations in stress spectra from that used in predictions; and, b) performing inspections at one-half the predicted time for the initial flaw to reach a critical size. Although the effect on the risks of structural failure due to these procedures is unknown, it should be noted that structural failures are relatively rare in comparison to other causes of accidents and class A and B mishaps [4].

To isolate the effect of NDI reliability on the risk of structural failure, consider first the nature of inspection uncertainty. Figure 2 presents a schematic of two POD functions. The ideal inspection system would detect all flaws larger than a_{NDE} and none smaller than a_{NDE}. This inspection system would have no uncertainty and would eliminate the inspection system from any calculation of probability of structural failure. No such system exists and no system approaching this ideal is anticipated in the near future. The realistic curve is typical of current inspection capabilities. Although the POD function approaches one as the crack length increases, a_{NDE} would have to be extremely large to obtain the very high POD values required for structural reliability, say, POD $(a_{NDE}) = 0.99999$ or greater. Note also that if a_{NDE} is determined on the basis of a fixed POD value, say $POD(a_{NDE}) = 0.9$, there is no information available which links a_{NDE} to the risk of failure. All that is known is that if there is a crack in the structure greater than a_{NDE} before the inspection, there is at least a 90 percent chance of detecting it. This is insufficient information to account for inspection uncertainty in calculations of failure probability.

Crack growth based calculations of the probability of failure during a usage period must account for the distribution of flaw sizes at the beginning of the period. Thus, the simplest link between inspection uncertainty and failure risk is the effect of the inspection on the flaw distribution that was present in the structure immediately prior to the inspection. In particular, let H(a) represent the probability of having a crack longer than a in the structure after the inspection (all detected cracks are repaired). Then

$$H(a) = \int_a^\infty [1-POD(x)]f(x)\,dx \qquad (1)$$

where

$1 - POD(x)$ = proportion of cracks of length x which are not detected.

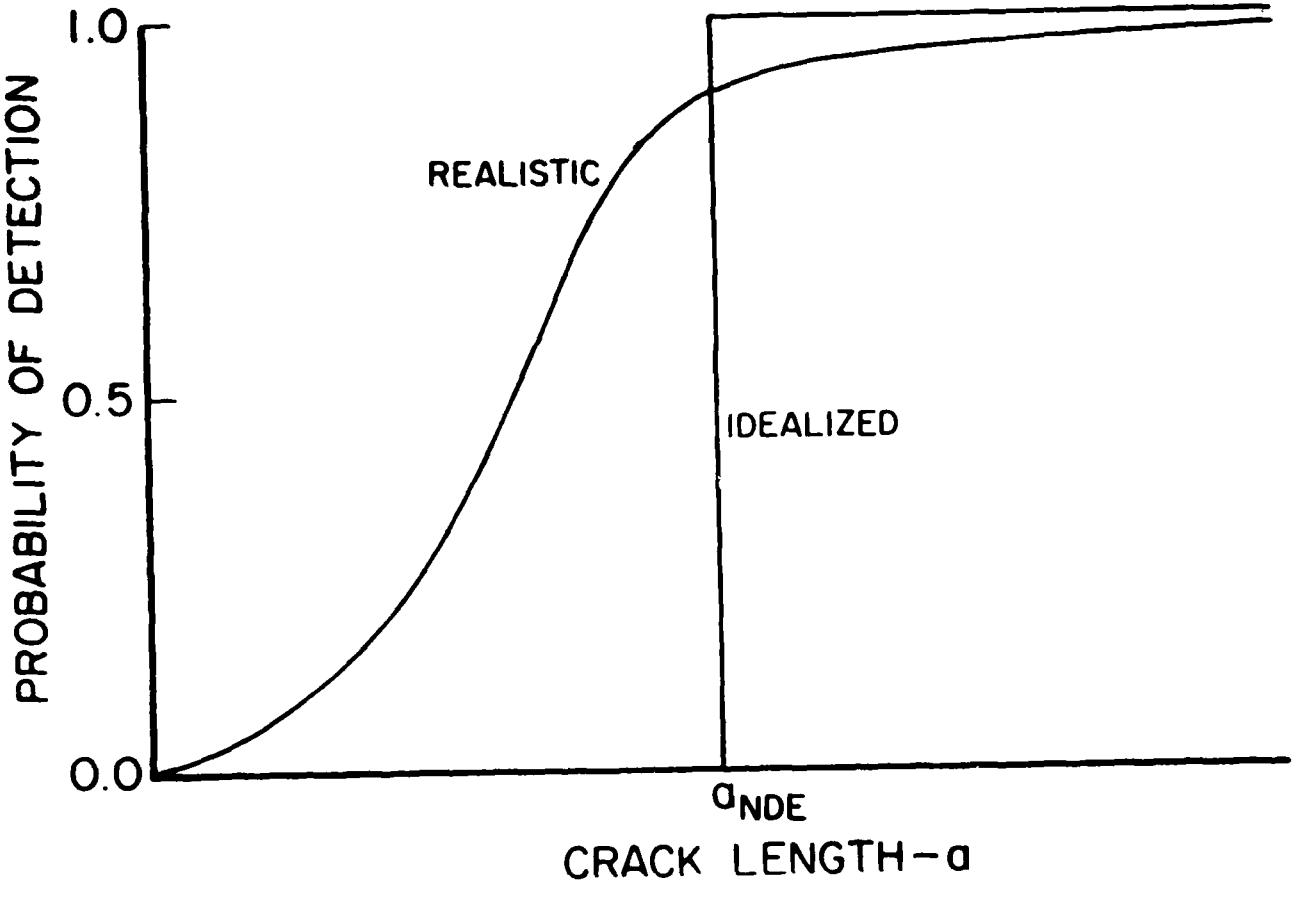


Figure 2. Schematic of Probability of Detection Curves.

$f(x)$ = probability density function of crack sizes in the population of details immediately prior to the inspection.

If it can be assumed that any crack longer than the crack length, a_{NDE}, will grow to critical size in the usage period and that no crack smaller than a_{NDE} will reach critical size, then $H(a_{NDE})$ represents the probability of failure during the interval. This assumption is not valid in the Air Force application of damage tolerance since periods of uninspected usage are scheduled at half the time required for a crack of length a_{NDE} to grow to critical. If this philosophy is conservative, $H(a_{NDE})$ is an upper bound on the probability of failure.

$H(a_{NDE})$ provides a measure of inspection efficacy for fixed a_{NDE} and known crack size distribution and POD function. Conversely, equation (1) can also be used to determine a crack length, say a^*_{NDE}, for which the probability of having cracks greater than a^*_{NDE} after inspection is less than a pre-specified probability q. a^*_{NDE} is defined by the equation

$$q = \int_{a^*_{NDE}}^{\infty} [1-POD(x)]\, f(x)\, dx = H(a^*_{NDE}) \qquad (2)$$

or

$$a^*_{NDE} = H^{-1}(q) \qquad (3)$$

Since H(a) measures the effect of inspection uncertainty in terms that would be required for any further crack growth based structural reliability studies, this function was selected as the basis for the evaluations of this program. The calculation of H(a) does require information about the crack size distribution prior to the inspection but this is the minimum additional information that could link inspection uncertainty to structural failure. Section 2.3 discusses the crack size assumptions used in this study.

There are several other calculations that will be used in the following. Again, let POD(x) represent the proportion of cracks of length x that are detected by a particular NDI system and f(x) represent the probability density function of crack sizes. The probability of inspecting a detail with a crack longer than a and detecting it is given by

$$G(a) = \int_a^\infty POD(x)\ f(x)\ dx \tag{4}$$

Note that G(0) is the total percentage of details for which a positive crack indication will be given, i.e., rejected by the inspection. Let F(a) represent the proportion of cracks with lengths less than or equal to a. Then

$$F(a) = \int_o^a f(x)\ dx \tag{5}$$

and 1-F(a) is the proportion of cracks greater than a. Since every crack greater than a is either detected or missed

$$H(a) + G(a) = 1-F(a) \tag{6}$$

Of particular interest is the proportion of cracks which are larger than a and are detected. This probability will be called the effective inspection reliability, ER(a), and is calculated by

$$ER(a) = \frac{G(a)}{1-F(a)} \tag{7}$$

The calculations of equations (1), (3), (4), and (7) will be used extensively in the evaluation of the sampling variability of the POD parameter estimates in Section 4 and in the interpretation of inspection efficacy for damage tolerance analyses of Section 5. The following paragraphs present the models used to describe the POD functions and the distributions of crack sizes used in the analytical studies of this program.

2.2 POD FUNCTIONS

The analytical model used to describe the POD function in this study was the log odds or logistic model as given by the equation

$$POD(a) = \{1+\exp - [\quad (\frac{\ln a - \quad}{\sigma})]\}^{-1}$$

The reasons for the choice of this model and method for estimating the parameters μ and σ are extensively discussed in Section 3. If $a_{0.5}$ represents the median crack detection capability, i.e. $POD(a_{0.5}) = 0.5$, then

$$\mu = \ln a_{0.5} \tag{9}$$

The parameter σ is a measure of the flatness of the POD function: the larger the value of σ, the slower the POD function approaches one.

Figure 3 presents the log odds POD function as obtained from 5 sets of NDI reliability data. The curve labeled $\sigma = 1.07$ has $a_{0.5} = 0.22$ in. and was obtained from eddy current surface scans of a C-130 wing box[1]. The curve labeled $\sigma = 2.06$ for which $a_{0.5} = 0.083$ in. was also obtained from eddy current surface scans but this time from C-130 wing panels. These are the AET and BET data sets, respectively, of Reference 1. The curves labeled $\sigma = 0.81$ ($a_{0.5} = 0.046$ in.) and $\sigma = 1.58$ ($a_{0.5} = 0.017$ in.) were calculated from the results of laboratory eddy current and ultrasonic inspection, respectively, of etched fatigue cracks in flat 2219-T87 Al plates (data sets 33 and 35 of Reference 5). The curve labeled $\sigma = 1.25$ ($a_{0.5} = 0.019$ in.) was calculated from inspection reliability data from an automatic eddy current bolt hole scan of fatigue cracked specimens. These data were obtained in a private communication from an Air Force Air Logistics Center. The curves of Figure 3 display the joint effects of $a_{0.5}$ (or equivalently, μ) and σ. The parameter μ tends to locate the curve at the shorter crack lengths while σ tends to dominate at the longer crack lengths.

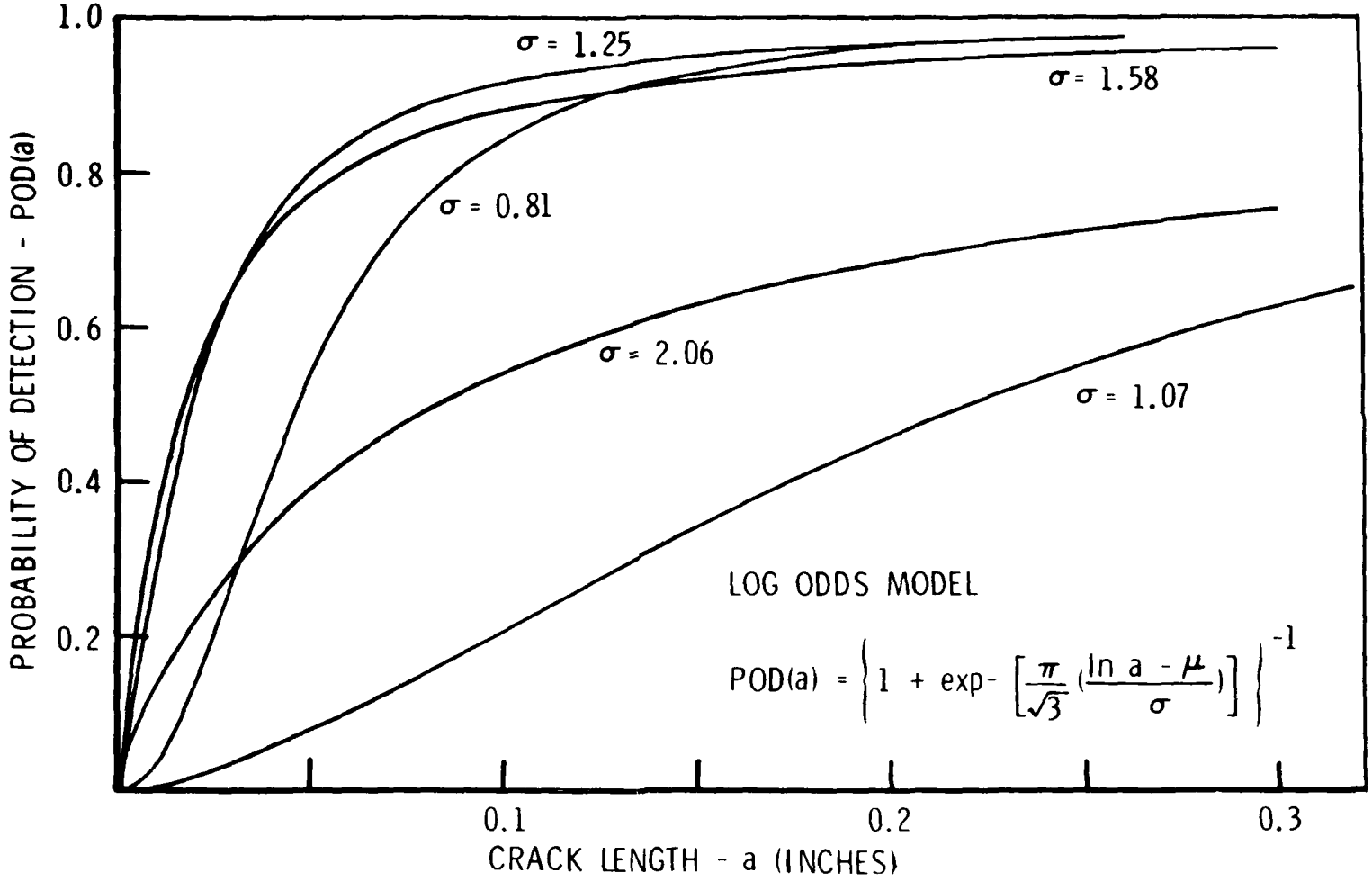


Figure 3. Log Odds POD Functions for Five Data Sets.

To eliminate one parameter from the sensitivity studies to be performed, all crack lengths were considered to be normalized by dividing by the mdeian crack detection capability, i.e.,

$$a' = a/a_{0.5} \tag{10}$$

For normalized crack lengths, $\mu = 0$ in the log odds POD model and the location and scale parameters of crack length distributions are expressed as multiples of the median detection capability.

Figure 4 displays POD as a function of normalized crack length for selected values of σ. As noted in Figure 3, existing manual NDI systems have σ values of about 1 or greater. Highly automated systems in laboratory reliability demonstration programs have produced σ values in the range 0.3 to 0.6. No current system has been shown to have a σ value as small as 0.1. Therefore, this pictured range of σ values was selected for the sensitivity evaluations of this study.

2.3 CRACK SIZE DISTRIBUTIONS

When a critical area of structure is inspected, there is a conceptual population of details each of which may contain a flaw. The population may arise from many locations on a single structure, from similar locations on many structures, or both. If an "unflawed" specimen is regarded as having a flaw of length zero (or extremely small), then a population of flaw sizes can be hypothesized for the inspection locations. Immediately prior to an inspection, the size of the particular "flaw" to be inspected is unknown beyond its distributional description.

The population of conceptual flaws at an inspection site can also be postulated without reference to a population of inspection sites. The material condition at an inspection location is the result of a number of random effects which produce flaws of varying magnitudes (including zero for no flaw). Examples of such random causes would be batch-to-batch production variation, the occurrence of voids or inclusions of various sizes, manufacturing or maintenance errors, or actual stress and environmental effects. Since the exact

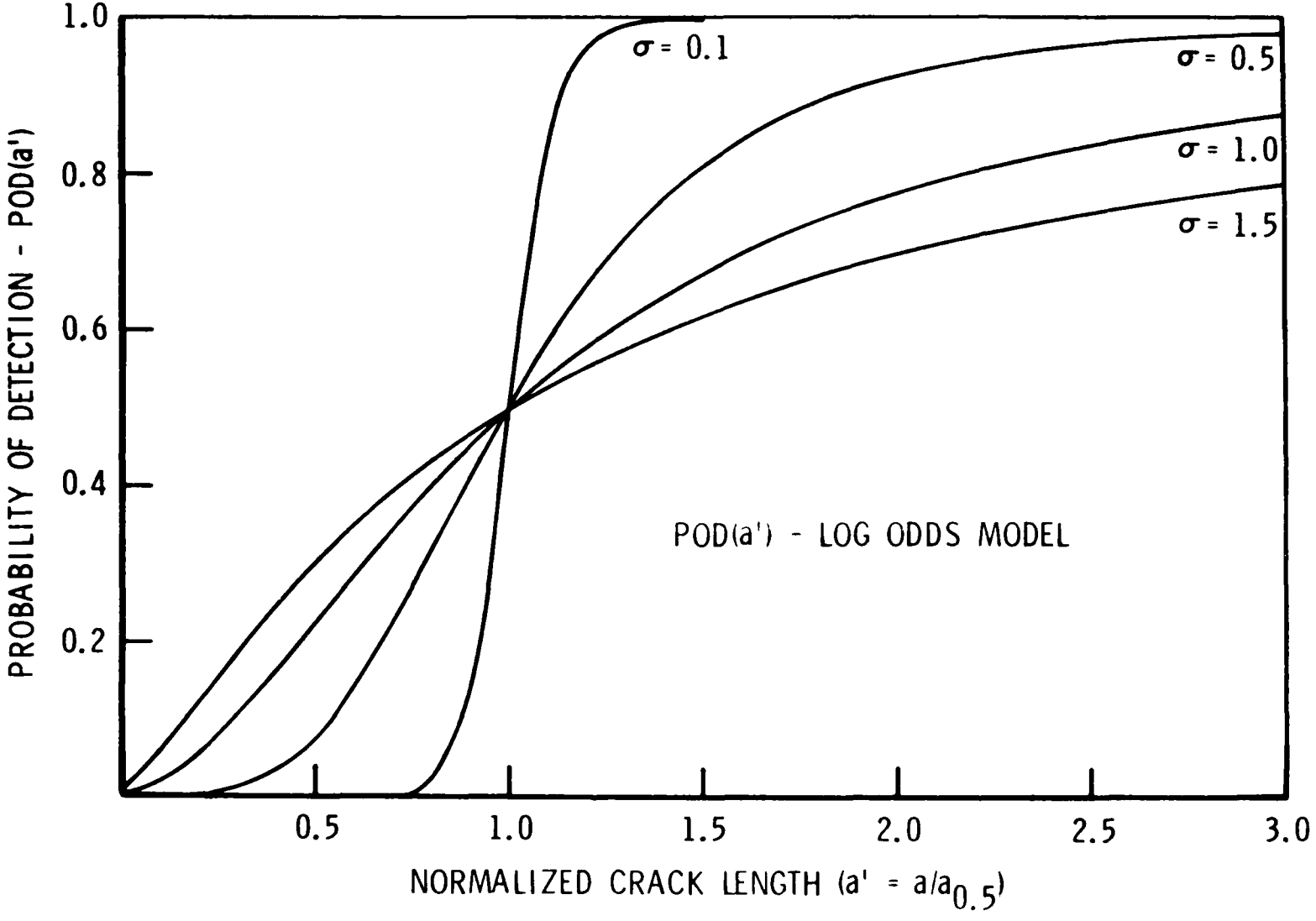


Figure 4. Example POD Functions for Normalized Crack Lengths.

Since the exact end result of these random stimuli is not predictable, the effects can be modeled in terms of a probability distribution which describes the potential flaw sizes at inspection.

The flaw size distribution is dynamic in a fatigue environment. At the start of the structure's life, the flaw size distribution is the result of the manufacturing and quality control process. As the structure is used, existing flaws will grow and new flaws will initiate and grow as a function of service time. Although this conceptual framework for the fatigue process is accepted, there is no general agreement on methods for modeling the process as there is very little data for estimating parameters and verifying probabilistic models of the existing approaches.

In this study, two approaches were taken to modeling flaw size distributions: the equivalent initial flaw size approach and a "rogue" flaw approach. These are discussed in the following paragraphs. In addition, a method for tracking a flaw size distribution throughout the life of a structure is also presented but the resulting complex characterizations of flaw size distributions were not used in the characterizations of inspection uncertainty.

2.3.1 Equivalent Initial Flaw Size Distribution

Starting with the F-4 damage tolerance assessment[6], and followed by the A-7 damage tolerance assessment,[7] the work on quantifying fastener hole quality of Potter[8], and recent durability studies,[9] the concept of an equivalent initial crack size distribution has been used to characterize the quality of structure. Very small initial "cracks" are equivalent to long times to crack initiation when subjected to standard crack growth models (even though, theoretically, the cracks are too small for the principles of fracture mechanics to hold). Although an artifact, this method of quantifying the initiation and growth of cracks is convenient in damage tolerant analyses. Since a crack growth life curve must be generated for every potential critical location, this analysis provides a model for deterministically "growing" the equivalent initial flaw size distribution. This approach assumes that the randomness in crack sizes at a particular point in the life of a structure is produced by the initial crack size distribution.

Different families of distributions have been employed as models for flaw size distributions. In Volume 4 of Reference 9. various families were tested against three sets of equivalent initial flaw size data and three families could not be rejected. Two of these were rather complex, 4 or 5 parameter families while the third was the Weibull family. For the purposes of this study the 2 parameter Weibull distribution was judged to be an adequate model for crack size distributions. The cumulative distribution function for the Weibull model is given by

$$F(a) = 1 - \exp\left[-\left(\frac{a}{\beta}\right)^{\alpha}\right] \qquad (11)$$

where β is the scale parameter and α is the shape parameter. In this study it was assumed that $\alpha = 1.5$ which is consistent with the data of Volume 4 Reference 9. The median crack length of the Weibull distribution is given by

$$\tilde{a} = \beta \exp\left[\frac{1}{\alpha} \ln \ln 2\right] \qquad (12)$$

or, for $\alpha = 1.5$,

$$\tilde{a} = 0.783\beta \qquad (13)$$

Thus, increasing β (or, equivalently, $\tilde{a}$) would correspond to the increasing crack sizes that would result in a fatigue environment.

In keeping with the normalized crack lengths of the POD functions, the scale parameters of the Weibull distributions were calculated to provide median crack lengths, $\tilde{a}'$, of 0.1, 0.25, 0.5, 0.75 and 1.0. Thus, median crack lengths were assumed to range from 10 percent of the median crack detection capability to the median crack detection capability. The crack size distribution for $\tilde{a}' = 0.1$ could be representative of initial quality in aircraft as a median normalized crack length of 0.1 could correspond to a real flaw on the order of 2 to 3 mils[7]. The distributions for large $\tilde{a}'$ values represent resulting damaging effects of fatigue crack growth and could be present at later stages in the structural life.

Figure 5 presents these crack size probability density functions as a function of normalized crack length. Also included in Figure 5 is the POD function with median detection capability of 1.0 and $\sigma = 1.0$.

2.3.2 Rogue Flaw Distribution

The second approach was directed at modeling the relatively large and relatively rare flaw which is not typical of the general population of structural details. These "rogue" flaws can be introduced in the structure as the result of a manufacturing error or accidentally during routine maintenance. Assuming that only a proportion, p, of all inspection sites contains a rogue flaw, p H(a) is the total probability of having a flaw greater than a and missing it at an inspection.

Three families of distributions were considered for modeling the rogue flaws. These were the uniform, triangular and exponential distributions. The uniform and triangular distributions are defined in terms of two parameters, the minimum and maximum possible crack lengths in the structural detail. The minimum value would be zero and the maximum would be a length, say A, that could not be exceeded. A realistic value for A would be the critical crack length, a_{cr}. The exponential distribution is defined in terms of one parameter, the average rogue flaw size, λ. Figure 6 displays the three candidate density functions where the exponential distribution is truncated at A.

Under the uniform distribution assumption, any size of a rogue flaw is equally likely in the possible range of lengths. The triangular distribution assumption implies that smaller values are more likely. The truncated exponential distribution has more small crack sizes for small values of λ and approaches the uniform distribution for large values of λ. In this study, A was chosen to be 10 or 20 in the normalized crack length scale.

The rogue flaw approach to modeling crack sizes has an intuitive appeal in that the inspection process is directed at the elimination of large cracks. In aircraft structure relatively

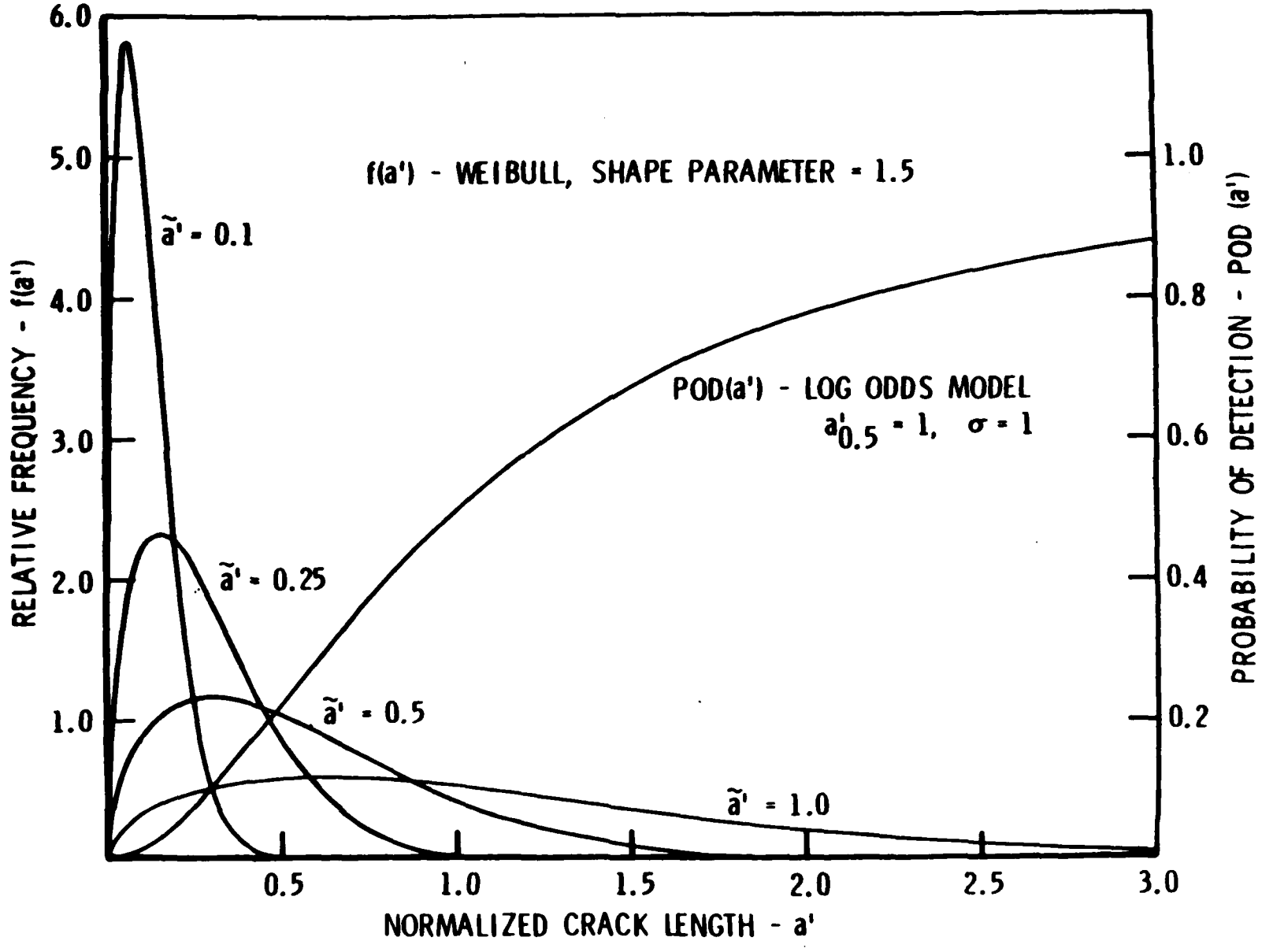
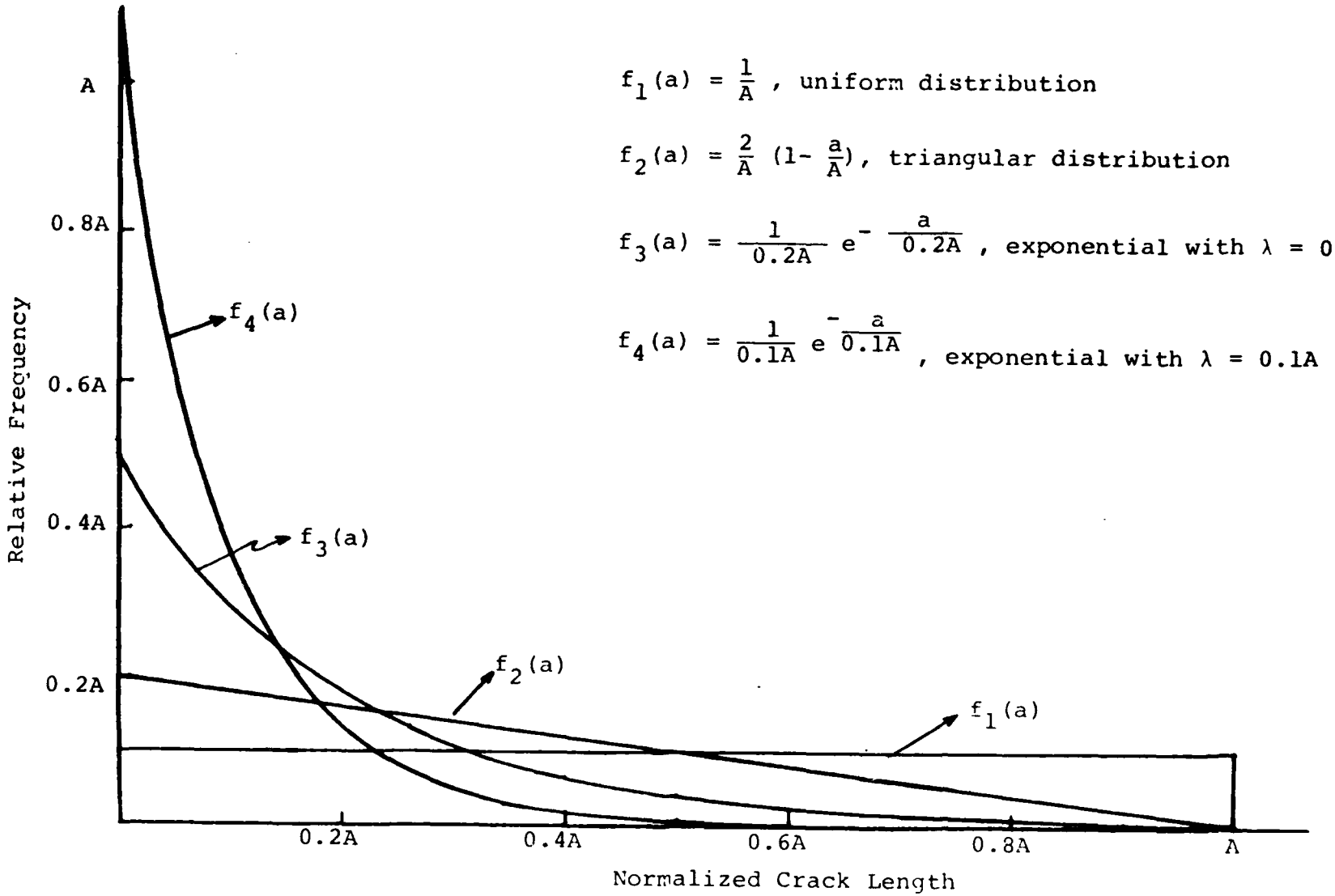


Figure 5. Crack Size Distributions as Modeled by Equivalent Initial Crack Sizes.



$f_1(a) = \frac{1}{A}$, uniform distribution

$f_2(a) = \frac{2}{A}\left(1 - \frac{a}{A}\right)$, triangular distribution

$f_3(a) = \frac{1}{0.2A} e^{-\frac{a}{0.2A}}$, exponential with $\lambda = 0.2A$

$f_4(a) = \frac{1}{0.1A} e^{-\frac{a}{0.1A}}$, exponential with $\lambda = 0.1A$

Figure 6. Example Rogue Flaw Probability Density Functions.

few cracks are detected in proportion to the total number of inspections. Thus, the cracks being detected can be viewed as the cracks in the upper tail of the equivalent crack size distributions. The shape of the density function in this upper tail region of the growing equivalent crack size distributions is reasonably close to the probability density functions being considered for the rogue flaw model.

The complete probability density function for flaw sizes under the rogue flaw model would be of the form

$$f(a) = (1-p)\ f_1(a) + p\ f_2\ (a) \qquad (14)$$

where p is the small proportion of details which contain "rogue" flaws whose density is $f_2(a)$ and $f_1(a)$ is the equivalent crack size density for modeling the initiation and growth of fatigue cracks as discussed previously. However, the mixture of distributions of equation (14) was not used in this study as the complexity of a four parameter model (at the minimum) was judged to be unwarranted. Rather, the equivalent crack size model, $f_1(a)$, and the rogue flaw model, $f_2(a)$ were treated separately and no attempt was made at combining the results.

2.3.3 Tracking Flaw Size Distributions

At an inspection, some of the structural details will be identified as containing flaws and these details will be repaired (in some sense). After the inspection, the population of details will be a mixture of the unrepaired and repaired structures. Given the pre-inspection crack size distribution, $f^-(x)$, the POD function, and the equivalent flaw size distribution for the repaired structure, $f_R(x)$, the post-inspection crack size density function, $f^+(x)$, can be calculated from the equation [10,11],

$$f^+(x) = G(o)\ f_R(x) + [1-POD(x)]\ f^-(x) \qquad (15)$$

where

$G(o)$ = total proportion of details for which a positive indication is given, equation (4).

This equation permits tracking the crack size distribution throughout the life of the structure by an iterative procedure:

a) Estimate the flaw size distribution for the structural elements prior to the manufacturer's quality inspection. Let $f_o^-(x)$ represent this density function.

b) Calculate the crack size density immediately after inspection, $f_o^+(x)$, by application of equation (12).

c) Analytically grow the cracks of the $f_o^+(x)$ under predicted service usage to the time of the first inspection. Let $f_1^-(x)$ represent the crack size distribution at this time.

d) Calculate the crack size density immediately after the first inspection, $f_1^+(x)$ by application of equation (12).

e) Continue steps c) and d) for the required number of inspections.

Although this process is analytically tractable, there are too many uncertainties in estimating the individual elements. The process is not practically useful at this time. Therefore, this complex model of flaw sizes was not used in this program. Rather, NDI reliability characterizations were evaluated on the basis of simple but plausible crack size distributions which could be present at each inspection and no attempt was made to predict the crack size distribution at the next inspection. This simplification did not affect the conclusions of the study.

SECTION 3
FLAW DETECTION RELIABILITY

Many factors influence the capability of an NDI system to correctly identify flaws in a structure. These include system factors which affect the ability of the system to consistently produce and interpret the information upon which flaw decisions are made. They also include factors which are characteristics of the individual flaws being inspected. The net effect is uncertainty in the detection of flaws so that the process of quantifying the inspection capability of a particular system requires a careful NDI reliability demonstration program coupled with a probabilistically based analysis of the data. This section presents an extensive discussion directed primarily to the analysis of NDI reliability data.

This section is organized into four parts. The first two discuss the definition of flaw detection reliability and the causes of uncertainty in an inspection process. The last two sections discuss the estimation of the probability of detection function including data analysis, sampling errors of parameter estimates and flaw size selection considerations in conducting an NDI reliability demonstration program.

3.1 DEFINITION OF FLAW DETECTION RELIABILITY

Flaw detection reliability is defined as the probability of detecting a flaw under pre-specified inspection conditions. As mentioned previously, this probability is associated with a crack length for applications in damage tolerance analysis.

Figure 7 shows a plot of inspection results on individual cracks emanating from fastener holes in a skin and stringer wing assembly as inspected by eddy current surface scans [1]. The points represent the proportion of times individual flaws were detected versus the length of the flaw. This figure illustrates

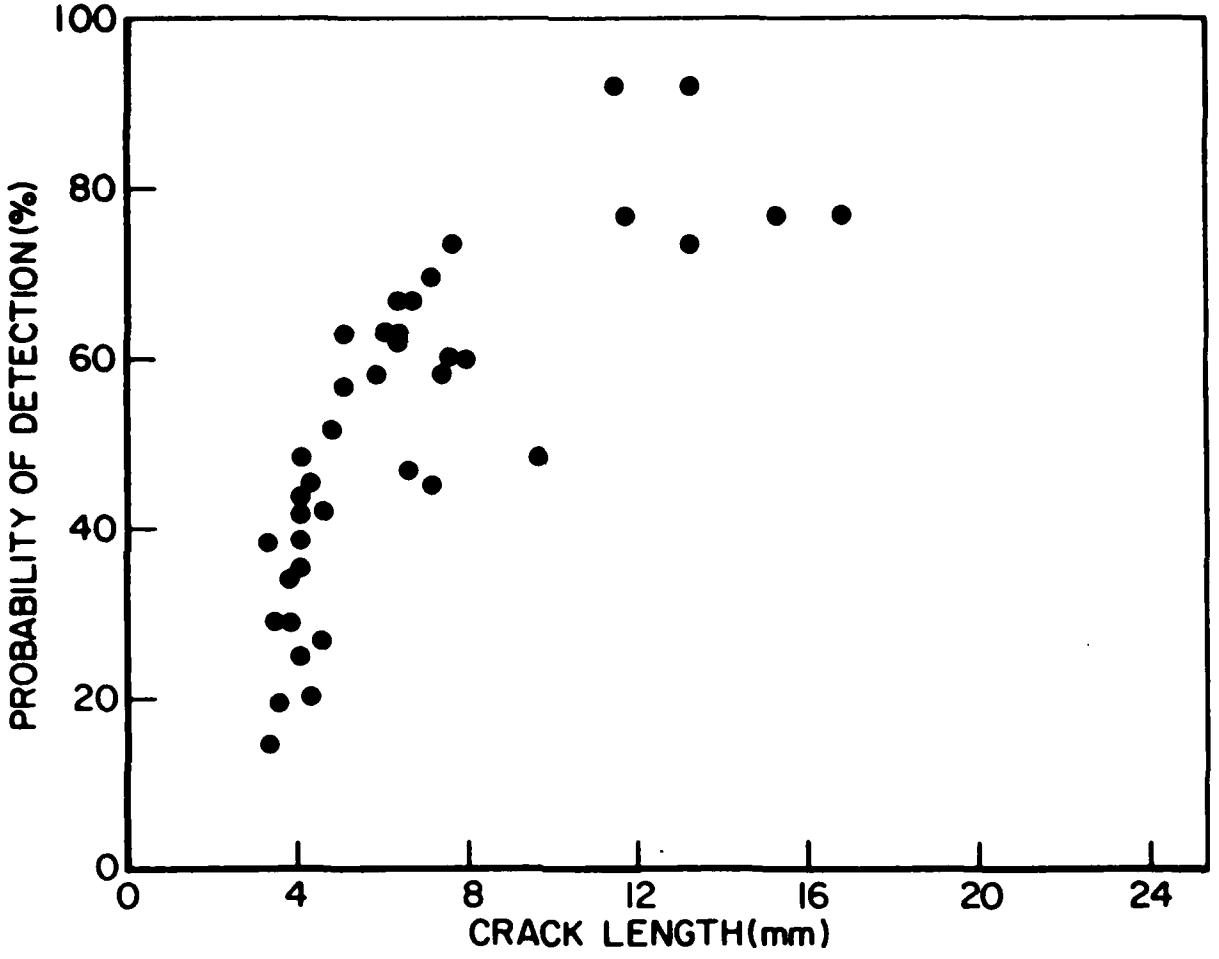


Figure 7. Crack Detection Probabilities for Individual Flaws as a Function of Crack Size.

that although the detection probabilities of individual flaws generally increase with crack length, not all flaws of the same length have the same detection probability. This variability in detection probabilities at a crack length has resulted in different interpretations of the definition of POD as a function of crack length.

The different interpretations center on the use of the mean trend in detection probabilities or a lower bound on detection probabilities as the method for characterizing the POD function. The exact formulation of the POD function should be consistent with reliability calculations for the structure under inspection. In calculating the reliability of a structure, the main concern is the proportion of flaws of a specific length that remain in the structure after an inspection. The POD function should, therefore, be defined as the proportion of flaws that will be detected as a function of crack length, i.e., the mean trend in detection probabilities as a function of crack length.

The distribution of detection probabilities at a crack length is illustrated in Figure 8. The curved line represents the general trend in detection probabilities as a function of crack length. The density function, $f_a(p)$, represents the distribution of detection of probabilities of all cracks of a length a. Let POD(a) represent the proportion of those cracks (i.e. of length a) that would be detected by the NDI system. $f_a(p)$ is the proportion of those cracks for which the detection probability is p. That is, 100 p percent of these cracks will be detected or 100 p $f_a(p)$ dp percent of all cracks will have a detection probability of p and be detected. Adding (integrating) over all values of p yields the total proportion of cracks of length a that will be detected. Therefore

$$POD(a) = \int_0^1 p\, f_a(p)\, dp \qquad (16)$$

The POD(a) function defined by equation (16) is the function which passes through the mean of detection probabilities at each crack length, i.e., the regression function.

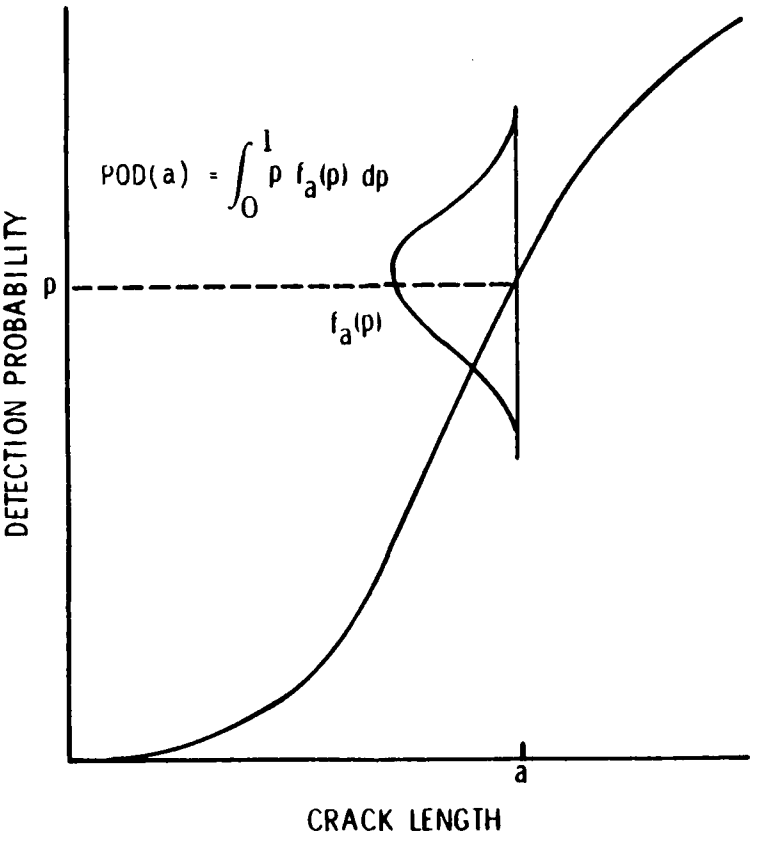


Figure 8. Schematic of Probability Density Function of Crack Detection Probabilities at a Crack Length.

Note that many individual cracks will have detection probabilities below the POD(a) value calculated by equation (16). This observation has led to attempts to characterize inspection capability in terms of some type of a lower bound on the detection probabilities at each crack length. This latter definition is different from that of traditional use, does not relate to the number of large cracks which may be missed at an inspection and will not be used in this study.

The variability of detection probabilities at a crack length does not influence the POD function; however, it does influence the method of analysing data to estimate the POD function. The following sections discuss the nature of the variability in detection probabilities and methods of analysing NDI reliability data collected with either single inspections per flaw or multiple inspections per flaw.

3.2 CAUSES OF UNCERTAINTY/SOURCES OF VARIATION

The reason for using sophisticated procedures in NDI is at the very heart of the uncertainty associated with the inspection results: the flaws that can cause failures are very hard to find. In defining what 'hard to find' means in relation to NDI, the nature of inspection uncertainty can be characterized. Although Figure 7 illustrates the concept of inspection uncertainty, it does not cover the whole inspection process. The POD is the measure of inspection uncertainty, not the cause. Causes of uncertainty are better defined in terms of the inspection process.

Typical NDI systems apply a stimulus to a suspect area and record the signal that returns from the specimen. A positive flaw indication occurs if the signal is higher than a threshold value. The flaws are hard to find for a number of reasons some of which are:

a) the operator does not know where the flaw is and therefore does not know where to aim the system,

b) material variability causes unpredictable changes in the stimulus before it reaches the flaw and in the signal before it returns to the NDI system,

c) variability in flaw geometry and orientation produces variability in the signal, and

d) calibration changes in the instruments from inspection to inspection reduces the predictability of the signal.

All of these inspection factors make flaws hard to find by contributing to the variability of the response signal; that is, the response signal in an inspection is a random variable. Since a flaw is detected if the response signal is larger than the threshold, the POD is the probability that the response signal is greater than the threshold.

In reasonable NDI systems the response signal increases with flaw size; however, as noted in the previous paragraph, the response signal and flaw size are not perfectly correlated. Figure 9 shows an example of the response signal as a function of crack length from a highly automated eddy current system as applied to fatigue cracks emanating from engine bolt holes (unpublished data from the General Electric Company). The response signal is measured as percent peak voltage and the crack lengths have been normalized to the length that results in a mean signal of twenty percent. In Figure 9, each dot represents the results of a single inspection and inspections made on the same flaw are connected by a straight line.

The variability in the response signal (á) is consistent with the difficulties associated with the inspection process; however, there is an interesting pattern to the deviations from the mean. The â values from a single flaw are typically grouped around a point that is shifted from the mean curve. This pattern of grouping indicates that there are two sources of variation in the response signal. One source is the variability in the mean â from flaw to flaw, and the other is the variability in â from inspection to inspection of the same flaw.

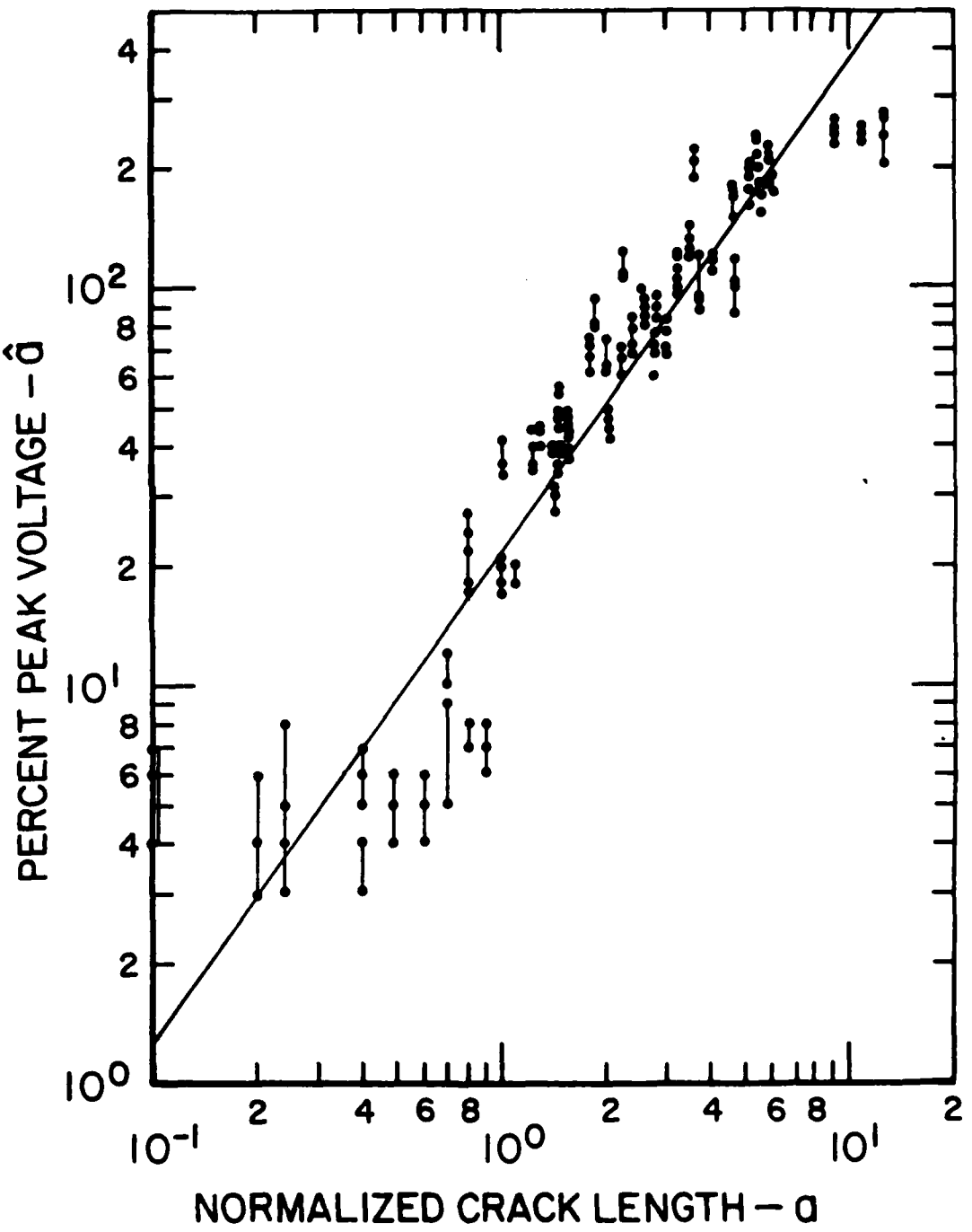


Figure 9. Observed Peak Eddy Current Response Voltages From Two Measurements on Each Crack.

The causes of uncertainty listed earlier can be grouped to correspond to these two sources of variation. The material properties, the flaw location, geometry, and orientation, and the pattern of residual stresses are strictly associated with individual flaws and do not change from inspection to inspection. Factors that do change from inspection to inspection include human factors such as attentiveness, skill, and health and equipment factors such as transducer variability and calibration.

Because of these two distinct sources of variation, the response signal has a compound distribution. This can be seen in the steps of an inspection process. First a flaw is picked at random along with its individual mean â. Then the human factors and equipment factors come into play resulting in a random deviation from the flaw mean for an individual inspection. These are two distinct random processes with distinct random variables.

The statistical procedure used to model compound distributions is called a variance components model. The variance components model breaks down the response signal into components that can be attributed to specific sources of variation. For â values the model would be:

$$\hat{a} = f(a) + c + e \qquad (17)$$

where f(a) represents the overall mean trend in â as a function of a, c represents the flaw to flaw variation, and e represents the variation from inspection to inspection of the same flaw. The function f(a) is fixed while the variables c and e are random with means of 0. The random variables c and e are referred to as the variance components of the model.

The importance of equation (17) comes into play in analyzing a set of data to estimate the reliability of an inspection system. All statistical analyses rely on assumptions about the underlying distribution of the response variable and usually all observations are assumed to be independent. Equation (17) points

out that when more than one inspection is made of an individual flaw, the results are not totally independent. Correlation between inspections of the same flaw occurs because all inspection of a single flaw have the same value for c. Therefore, specialized statistical procedures are required when analysing NDI reliability data that includes multiple inspections of individual flaws. These methods will be presented in Subsection 3.3.3.

3.3 DATA ANALYSIS

The data collected in NDI reliability demonstration programs have an unusual probabilistic structure and specialized analysis techniques, as well as standard techniques applied to transformations of the variables, are required. This section discusses the analysis of NDI reliability data, starting with a discussion of the validity of the log logistic function as the basis of the analysis. The last two sections present specific analysis methods for data in which only the inspection result (pass or fail) is recorded and for data in which the response signal ($\hat{a}$) is recorded.

3.3.1 Models for POD Functions

One of the more controversial aspects of NDI reliability estimation is the selection of a model for the POD function. In fact, the first attempts at estimating POD did not assume any model. The crack lengths used in the first studies were divided into subintervals and binomial distribution statistics were used to put confidence bounds on the POD for each interval. This often resulted in erratic curves for the lower confidence bound because the sample size changed from interval to interval. Attempts at smoothing the lower confidence curve involved intervals with fixed numbers of cracks and overlapping interfals; however, these approaches were extremely conservative because the confidence bound for an interval was plotted at the upper endpoint of the interval. Another problem with the early approaches was that the lower confidence bounds on the POD could not approach one.

The use of an assumed model for POD greatly simplifies the task of estimating NDI reliability. POD estimation with an assumed model reduces to estimating the parameters of the model rather than the POD at crack length intervals. The smoothness of the POD curve and its confidence bound is supplied by the model rather than an artificial interpolation technique and the confidence bound can be formulated so that it converges to one.

Once the decision has been made to use a model for POD estimation, the only question is what form the model should take. A previous study [2] suggested that the log logistic (or log odds) function was a suitable choice based on the data from the "Have Cracks Will Travel" program[1]. The log logistic function was picked over the Weibull distribution, the inverse Weibull distribution, the log normal distribution, and the arcsine distribution based on goodness of fit of the mean trend and on the structure of the deviations from the mean. The log logistic function provided a good fit to the "Have Cracks" data over a wide range of conditions; however, the evidence is still limited to this one study.

An analysis of the data collected by Yee et al.[5] was conducted to further investigate the general validity of the log logistic model assumption. Eight data sets were analysed with essentially the same results. The analyses of two of the sets of data are presented in Figures 10 and 11. Each figure contains plots of POD curves for NDI reliability data collected by a single operator using either ultrasonic (Figure 10) or eddy current (Figure 11) equipment. The three curves in each plot represent the results of analyses using the Weibull (W), the log-logistic (L.O.) and the lognormal (L.N.) distribution functions.

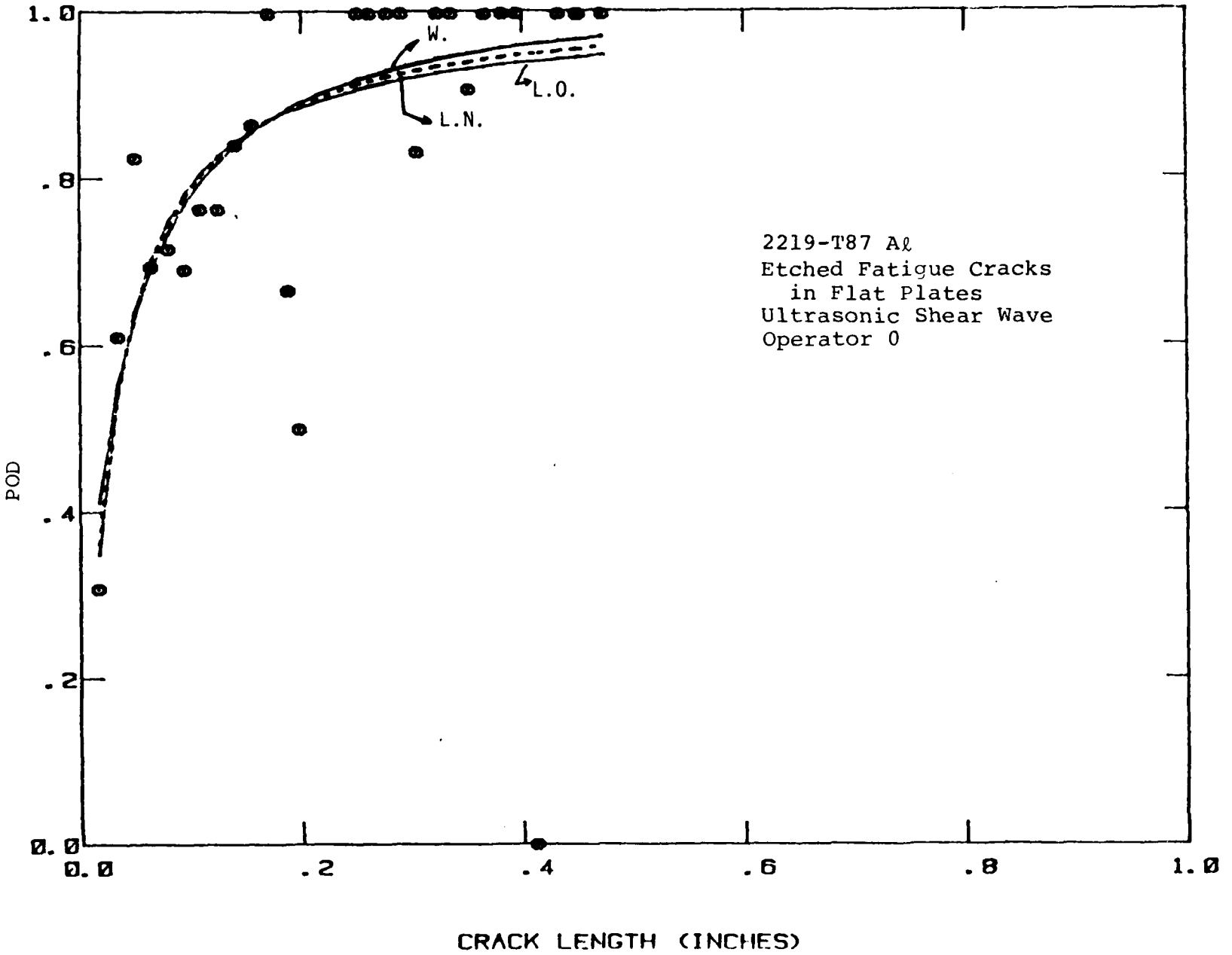


Figure 10. Comparative Analysis of Ultrasonic Inspection Capability.

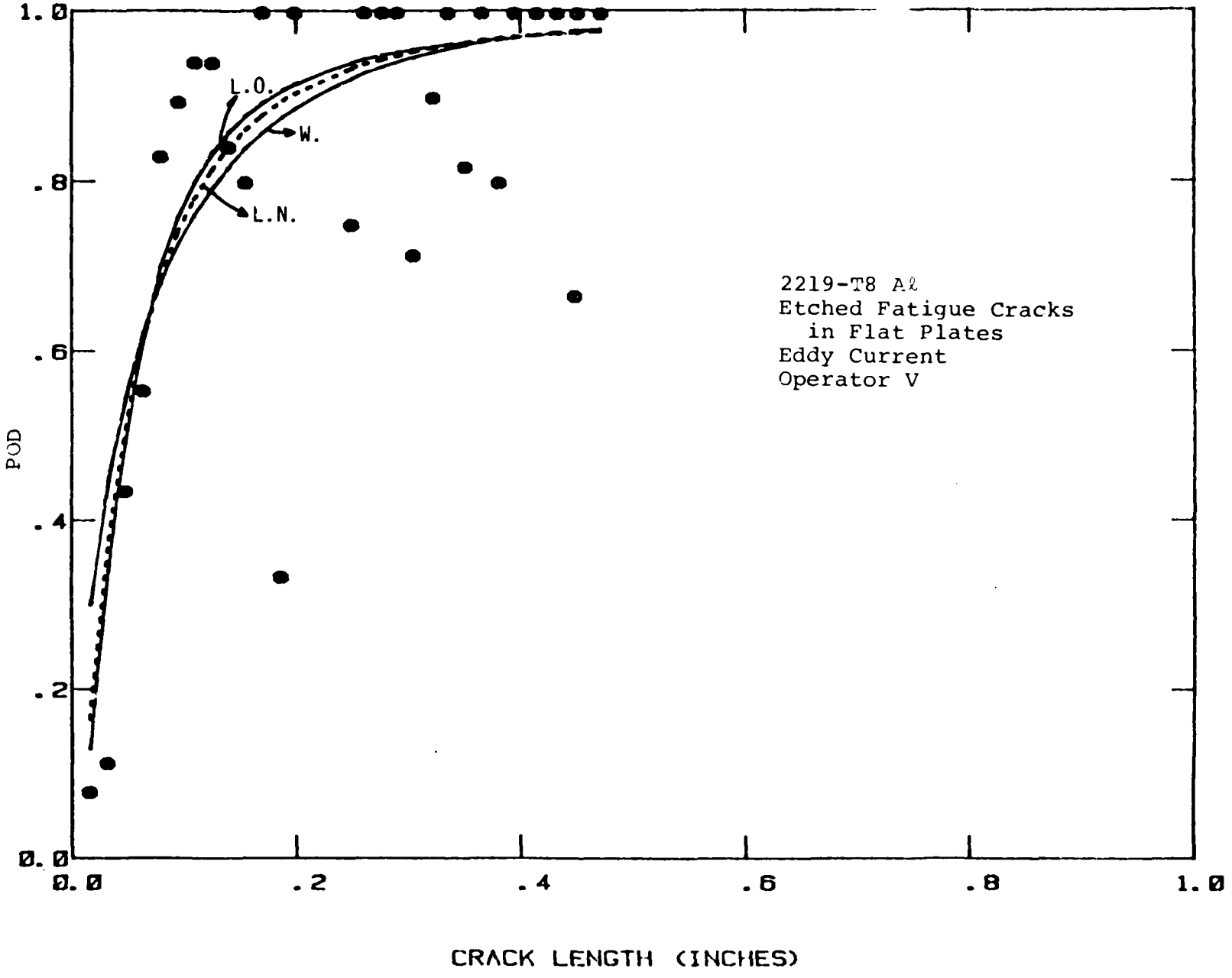


Figure 11. Comparative Analysis of Eddy Current Inspection Capability.

In general, there is very little difference between the fit provided by the three models; however, there is a systematic pattern to the three curves. In each case the Weibull distribution function provided the highest estimate of POD in the tails of the function while the log logistic distribution function always provided the lowest estimate of POD. This indicates that the log logistic function represents a conservative choice for the POD model since structural safety is concerned more with the POD at large crack lengths than with the POD at short and medium crack lengths.

A further illustration of this concept is seen in Figures 12-14. In these Figures, Weibull and log logistic functions were forced to pass through a common pair of points. Since these are each two parameter functions, they will only meet in two points. Figures 12 and 13 show a log logistic function and two Weibull functions, one that meets the log logistic function at the fifty percent POD point and the ninety percent POD point and another that meets the log logistic function at the ten and fifty percent POD points. Figure 12 has a broad scatter in the POD function and Figure 13 has a short scatter. Figure 14 shows Weibull and log logistic functions that meet at the ten and ninety percent POD points for a broad scatter and a short scatter.

The same ordering seen in the data collected by Yee et al. [5] holds in the curves of Figures 12-14 also. At crack lengths longer than the higher point of intersection and shorter than the lower point of intersection, the Weibull function is higher than the log logistic function. This indicates that an analysis based on the Weibull distribution function would tend to over estimate the probability of detection for long cracks when the assumptions are not met. The log logistic function, on the other hand, provides a degree of conservatism when used as the basis for the analysis of NDI reliability data.

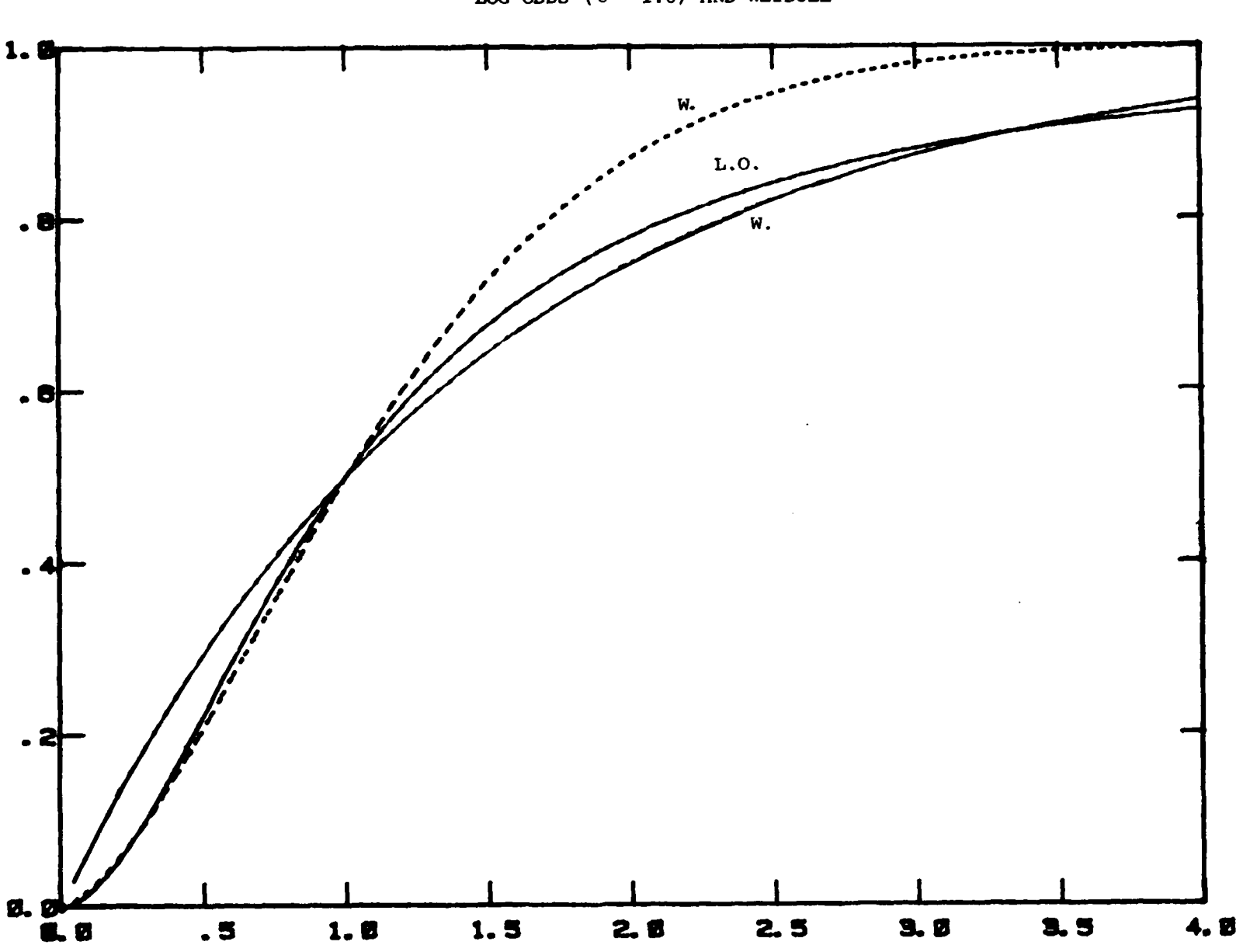


Figure 12. Comparison of a Log Logistic Function with Similar Weibull Functions (broad scatter).

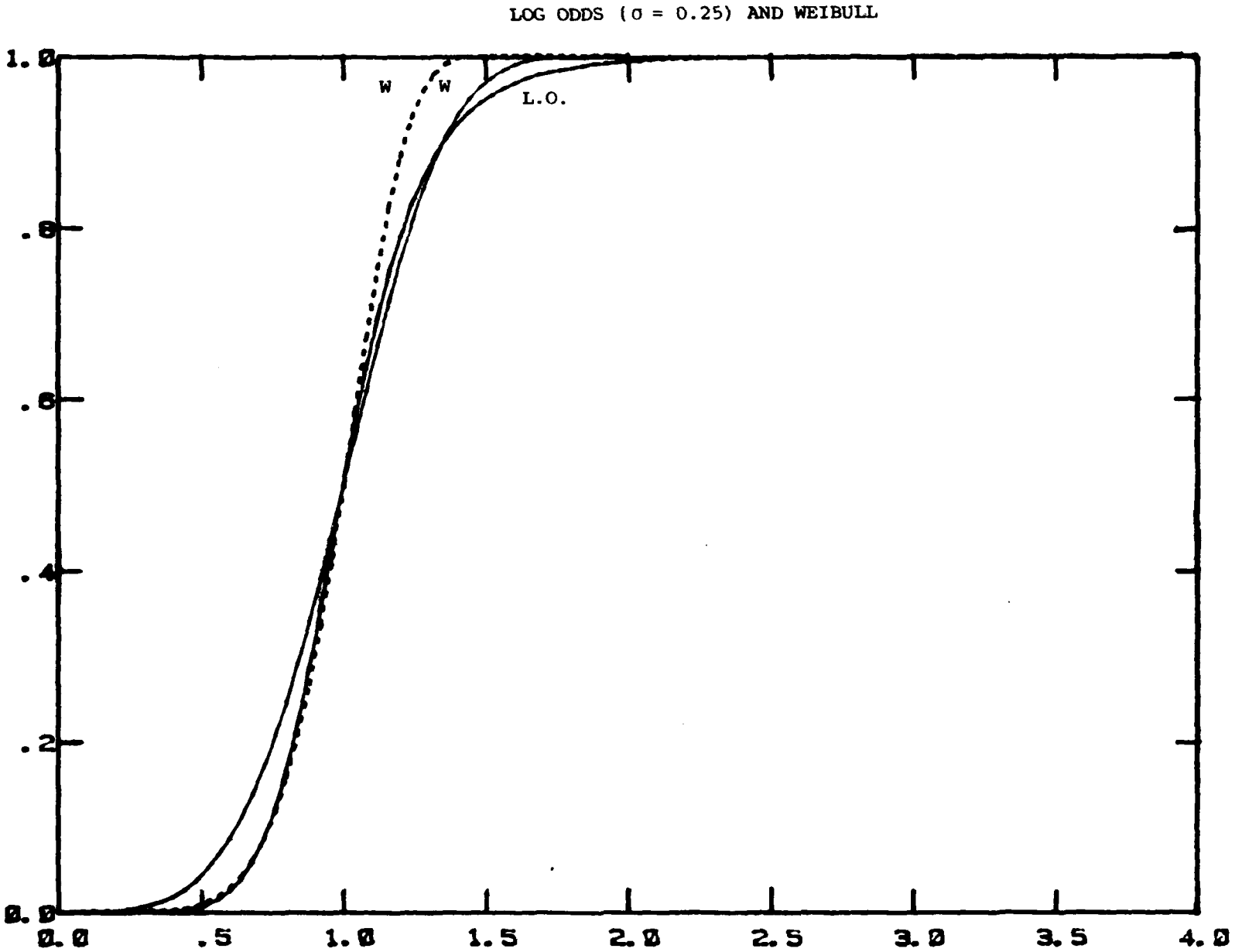


Figure 13. Comparison of a Log Logistic Function with Similar Weibull Functions (short scatter).

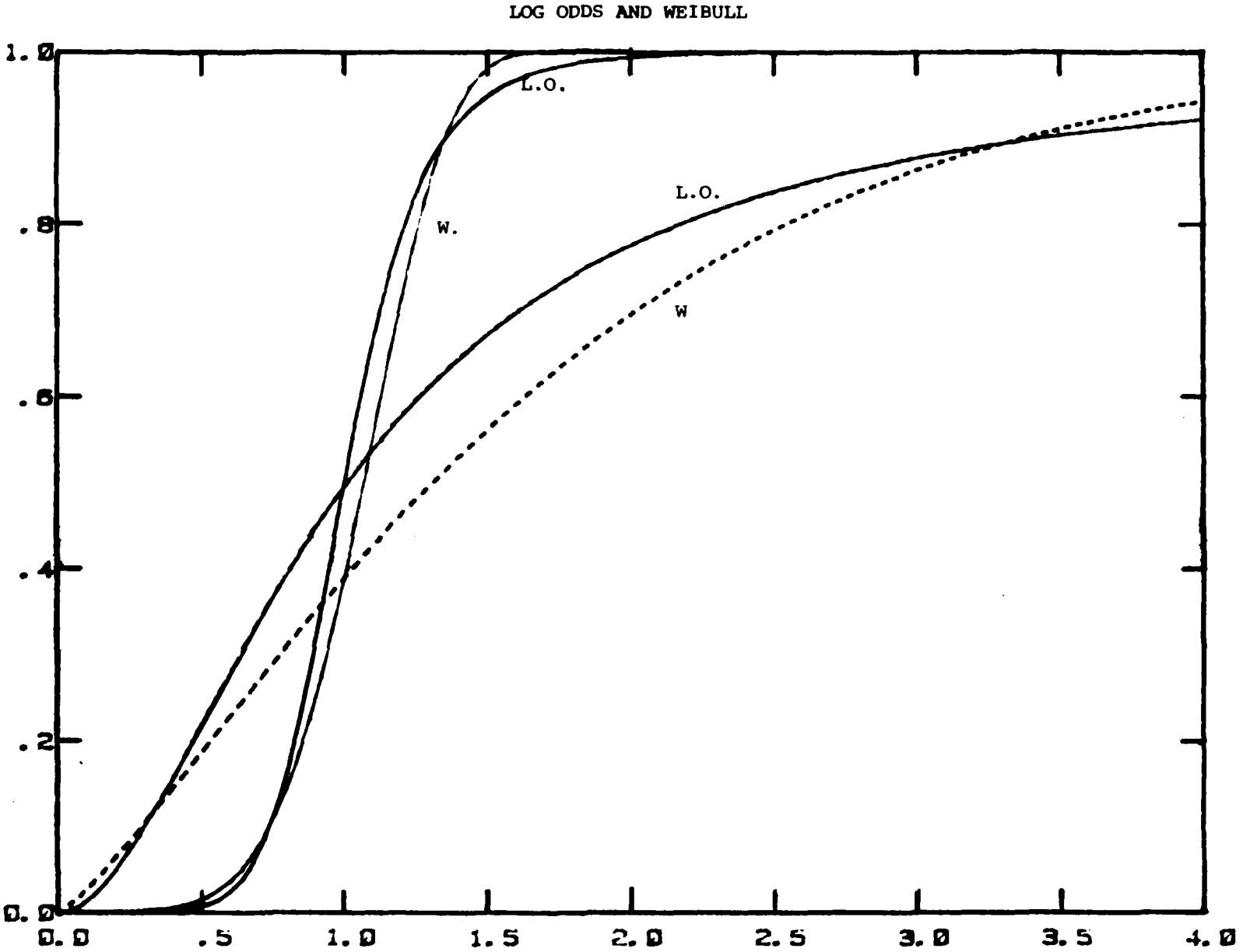


Figure 14. Comparison of a Log Logistic Function with a Similar Weibull Function for Broad and Short Scatter.

The log logistic function has a number of other features that make it a good choice for use as the model in analysing NDI reliability data. It is analytically simple and results in estimates that are similar to estimates provided by the Weibull and log normal distribution functions. In fact, the log logistic function is a close approximation to the cumulative log normal function. In addition, the analysis based on the log logistic function results in conservative estimates of POD in the safety critical areas of the POD function. Therefore, the log logistics (or log odds) model was selected as the POD model for all of the remaining analyses of this program.

3.3.2 Pass/Fail Data

Historically, NDI reliability data has been collected as a crack length along with an indication of whether or not the flaw was found on a particular inspection. The crack lengths are determined through an independent means such as replicates or tear-down inspection. During the inspections the flaws are only identified by the inspection locations and the inspectors record whether each site passed or failed the inspection. Because most of the NDI reliability data currently available is in this pass/fail format, the analysis of pass/fail data is covered first.

There are two analytically distinct ways to run an NDI reliability demonstration experiment: one inspection per flaw or multiple inspections per flaw. For data collected with a single inspection per flaw, all the observations are independent and the analysis is reasonably simple. Multiple inspections conducted on the same flaw will be correlated so that there are dependencies between observations when more than one inspection is made on each flaw. These two types of experiments require different analyses and are covered in separate sections.

3.3.2.1 One Inspection Per Flaw

There are two types of analyses based on the log logistic model that can be applied to single inspection pass/fail data. If the data can be conveniently grouped by crack length, a regression analysis can be performed on transformed variables. Typically, however, the crack lengths used in

an NDI reliability demonstration program cannot be conveniently grouped. Maximum likelihood methods utilize the individual inspection results in the (0,1) form and can therefore be applied without grouping the data. Regression methods are more appropriate for multiple inspection per flaw data and will be discussed in the next section. This section deals with maximum likelihood estimates of POD as modeled by the log logistic model.

The mathematical details of maximum likelihood analysis of single inspection per flaw NDI reliability data are covered in Appendix A, Section A1.2. The log logistic model is reparameterized to provide a linear form in transformed variables and to provide simpler maximum likelihood equations. The maximum likelihood equations (A18) and (A19) are solved simultaneously for the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\beta}$. Estimates of the equation (8) location and scale parameters, μ and σ are then given by:

$$\hat{\mu} = -\hat{\alpha}/\hat{\beta}$$
$$\hat{\sigma} = \pi/(\hat{\beta}\sqrt{3}) \qquad (18)$$

The estimates of POD and its lower confidence bound are calculated using the linearized form of the log logistic model and the parameter estimates $\hat{\alpha}$ and $\hat{\beta}$.

Figures 15-17 contain some examples of maximum likelihood estimates (MLE) of POD from single inspection per flaw NDI reliability data. The data were collected by the General Electric Corporation under a contract with the Air Force Materials Laboratory and are presented in a normalized scale based on the crack length at which the POD was 50 percent. The arrows on the bottom and top axes represent inspection results: detections on the top axis and misses on the bottom axis. The solid line is the mean POD and the dashed line is the lower 95 percent confidence bound.

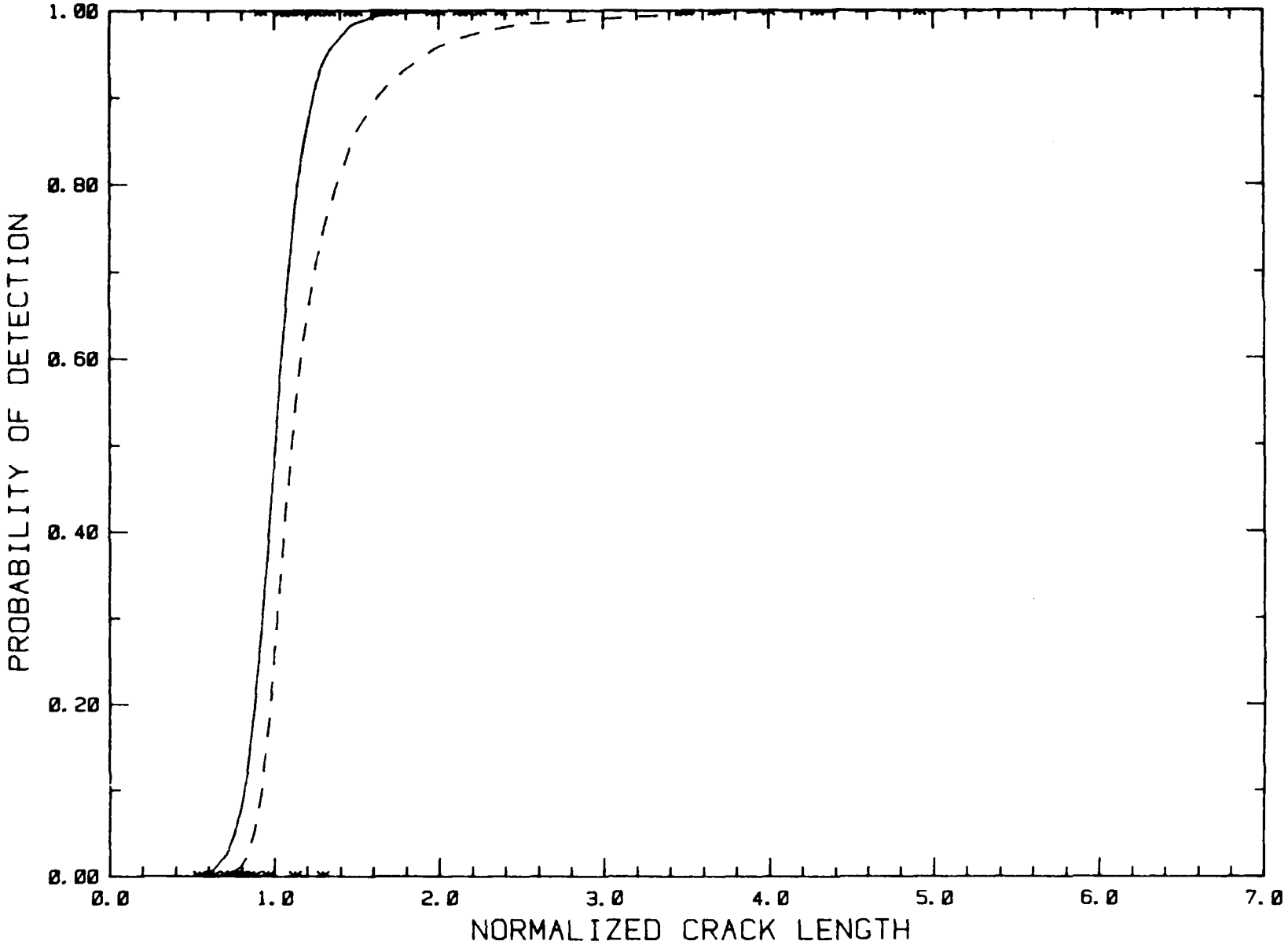


Figure 15. Normalized Example of MLE's for an Eddy Current Inspection Process.

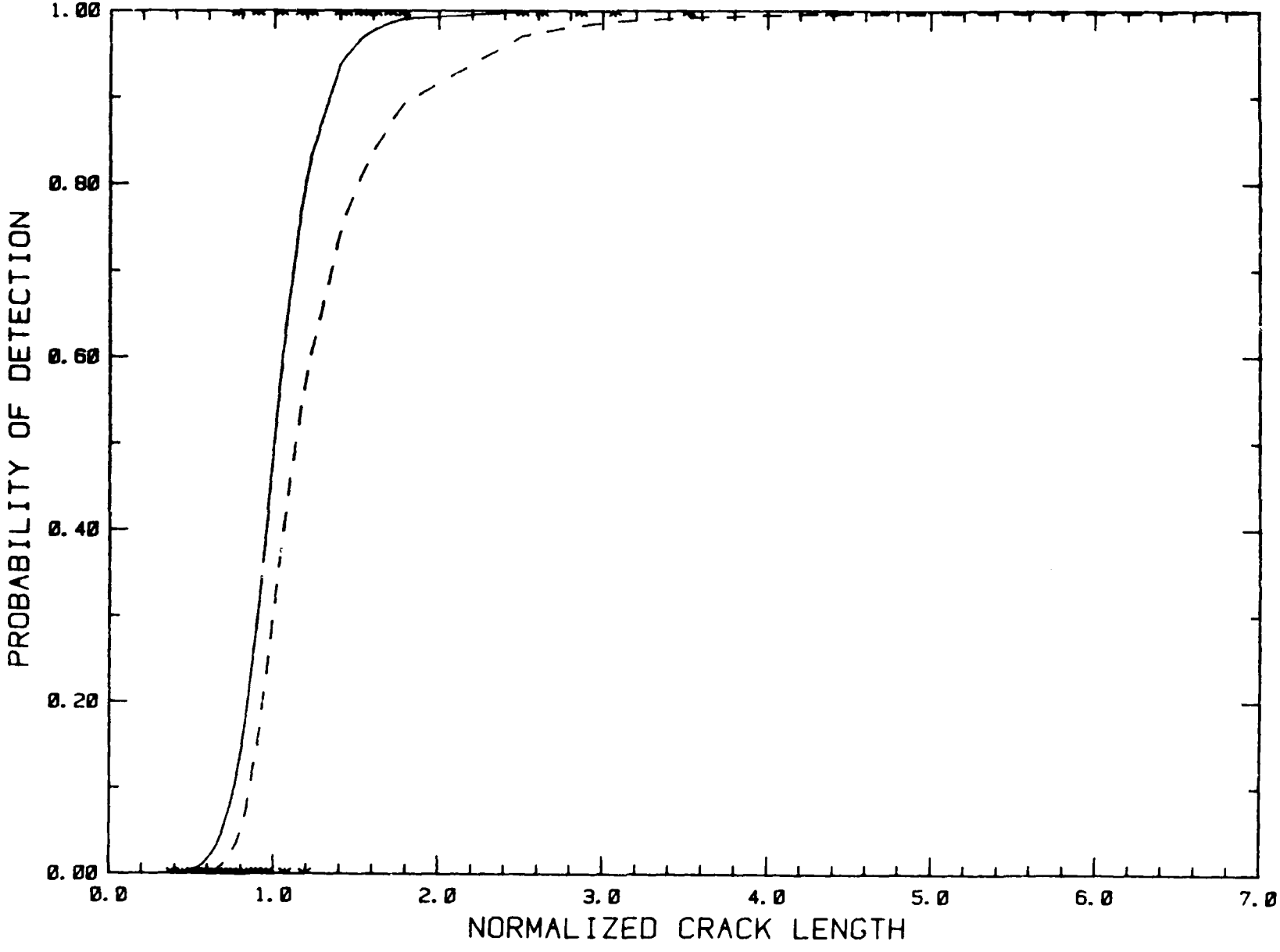


Figure 16. Normalized Example of MLE's for an Ultrasonic Inspection Process.

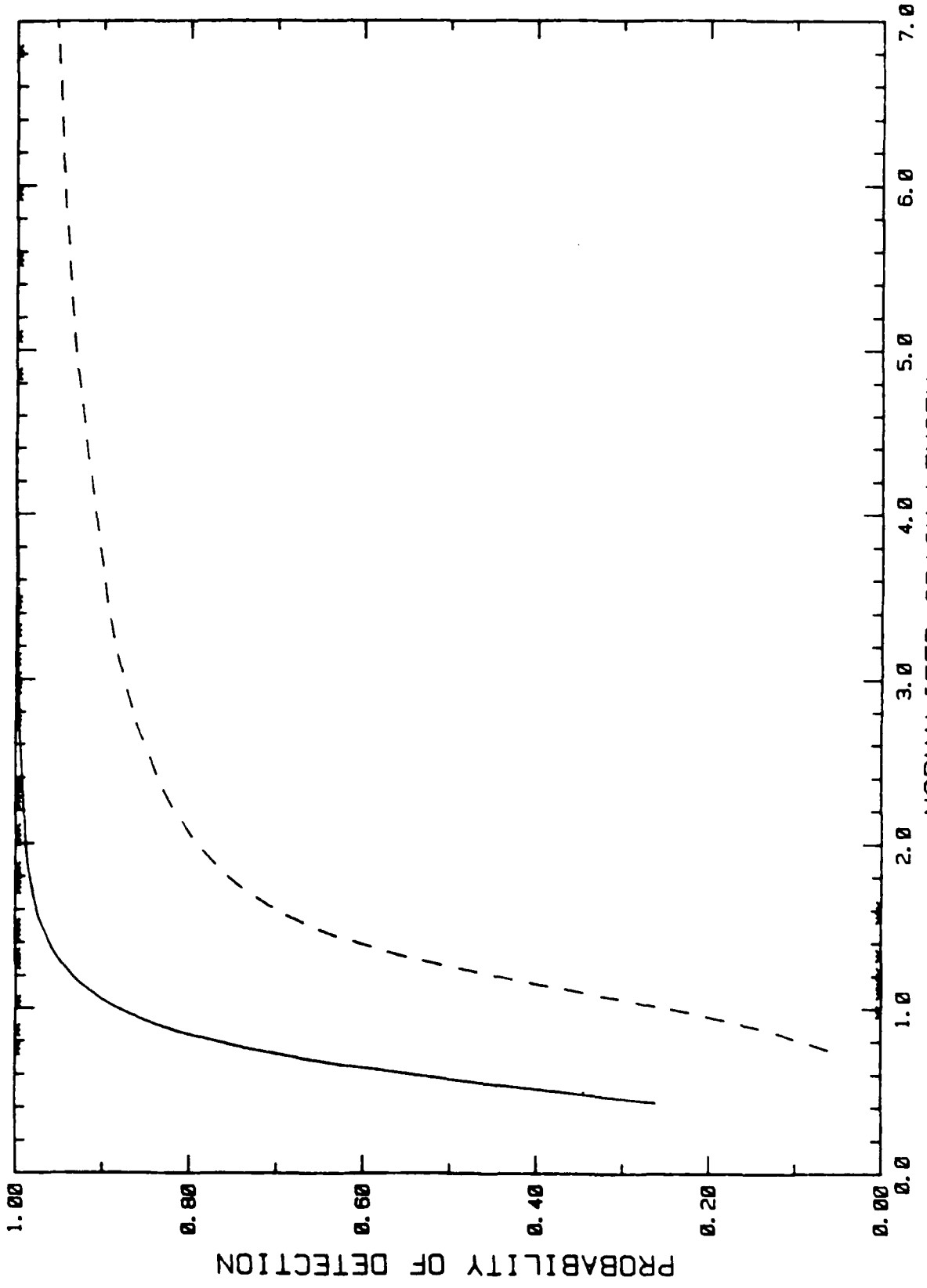


Figure 17. Normalized Example of MLE's for a Flouropenetrant Inspection Process.

These three curves were picked to represent three different capabilities in terms of the steepness of the POD function. Figure 15 shows the results for eddy current inspections and $\hat{\sigma} = 0.17$. Ultrasonic inspections are presented in Figure 16 with $\hat{\sigma} = 0.23$. The slowest ascent is seen in the fluorescent penetrant inspections shown in Figure 17 where $\hat{\sigma} = 0.50$.

Some caution is necessary in applying the maximum likelihood techniques of Appendix A. Iterative algorithms on a computer are required to solve equations (A18) and (A19) and there is a tendency to readily accept numbers produced by a computer. In some instances, the algorithm will work and estimates of the POD with a lower confidence bound will be produced; however, the estimate of β may not be significantly different from 0, rendering the mean curve and the confidence bound meaningless.

Figure 18 shows an example of a nonsignificant estimate of β. There is only one miss in the data of Figure 18 so that there is no trend in POD as a function of crack length. The computer algorithm still computed estimates of the POD and its lower confidence bound; however, the lower confidence is not monotonically increasing. When β is significantly different from 0, the lower confidence will converge to 1. An estimate of β that is not significantly different from 0 basically means that the data do not display a trend in POD as a function of crack length: that is, the POD is the same no matter how long the crack.

Another problem results from the fact that there might be more than one solution to equations (A18 and A19). The particular solution that the computer algorithm will converge to is a function of the initial estimates that are used. A method that produced good initial estimates in analysing the GE data is the method of moments. Method of moments estimators are linear combinations of the inspection results and are therefore easy to calculate. The details of method of moments estimators are given in Appendix A.

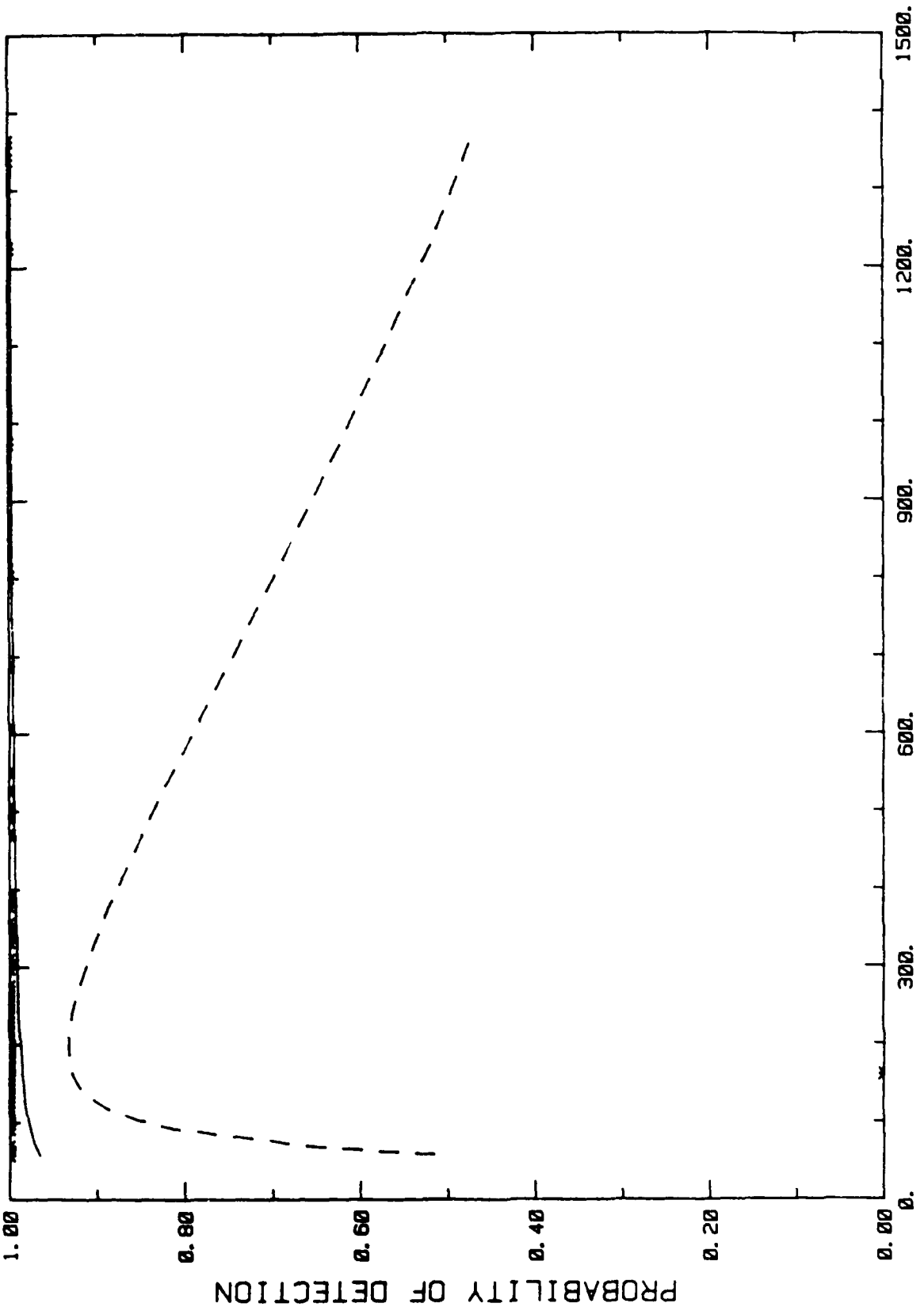


Figure 18. Normalized Example of a Nonsignificant POD Fit.

3.3.2.2 Multiple Inspections Per Flaw

There are two sources of variation in inspection results when more than one inspection is performed on each flaw. First, the detection probability for an individual flaw is a random variable. Then, since detection probabilities are not exactly 0 or 1, there is random variation in the number of times an individual flaw is detected. Both maximum likelihood and regression analysis methods can be used to analyze these types of data. These methods are discussed in the following sections.

3.3.2.2.1 Maximum Likelihood Estimates

Maximum likelihood estimates are based on the probability density function of the observed random variable; in this case, the number of detections for each flaw. For single inspection data, the number of detections for a single flaw is a Bernoulli random variable with probability of success given by POD(a). The distribution of the number of detections for a single flaw becomes more complicated when more than one inspection is performed. The conditional distribution for the number of detections for a single flaw given the detection probability is a binomial random variable. The unconditional distribution is not so easily described. The unconditional distribution involves the integral of binomial probabilities times the density of detection probabilities.

There are a large number of potential probability models for the density function of detection probabilities. The use of most of them, however, does not result in a closed form for the unconditional distribution of the number of detections of a single flaw. One exception is the Beta distribution and the resulting unconditional distribution of the number of detections is called a beta binomial distribution. Williams [12] shows how

the beta binomial distribution can be used to estimate the parameters of a log logistic function.

A modification of Williams method was used to reanalyse the "Have Cracks Will Travel' data and some of the results are presented in Figures 19-21. In all three figures the plain solid line is the estimate of the POD function from a regression analysis, the solid line with vertical bars is the estimate of the POD function from the maximum likelihood analysis, and the dashed line represents the POD minus two detection probability standard deviations calculated with maximum likelihood estimates.

In all three figures, the POD curves from the two methods are approximately equal and in Figures 19 and 20 the dashed line bounds the approximately correct proportion of flaws; however, in Figure 21, the lower bound falls on top of the POD curve even though the detection probabilities for about 50 percent of the flaws fall below it. This highlights a problem with maximum likelihood anslysis for multiple inspection data when the number of inspections per flaw is not sufficiently large.

The data in Figures 19 and 20 had 62 and 54 inspections per flaw, respectively. The detection probability for each flaw is estimated fairly precisely and therefore the variability in detection probabilities can be easily estimated. The data in Figure 21 had 9 inspections per flaw. The variability associated with estimating the individual detection probabilities for the data in Figure 21 is large enough to mask the variability in detection probabilities from flaw to flaw. The maximum likelihood estimates will converge to 0 or a negative number (usually near 0) for the estimate of the flaw to flaw variation in detection probabilities in the type of situation illustrated by Figure 21.

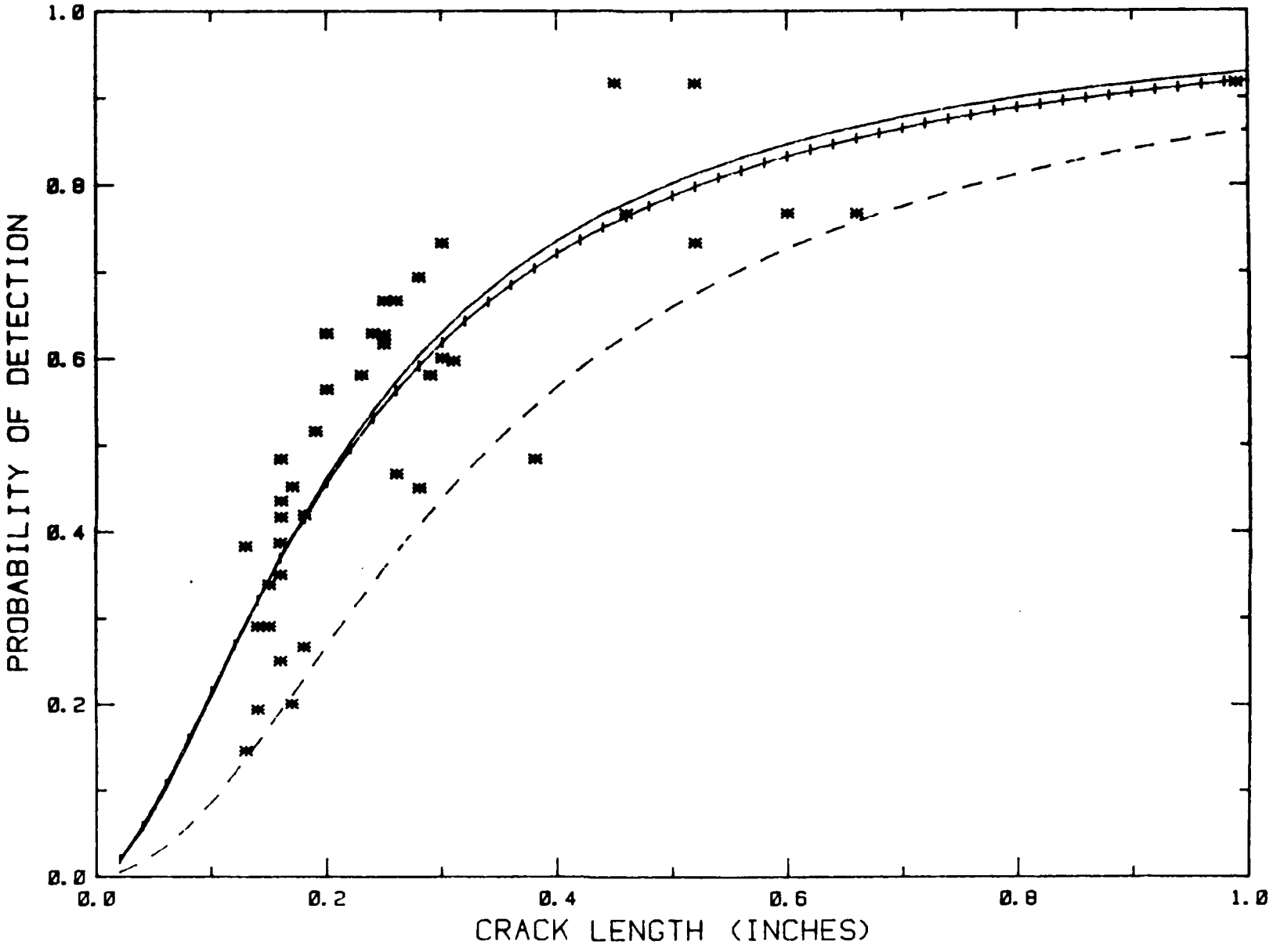


Figure 19. Comparison of Regression and Maximum Likelihood Estimates of POD for Eddy Current Inspections of a C-130 Skin and Stringer Wing Assembly.

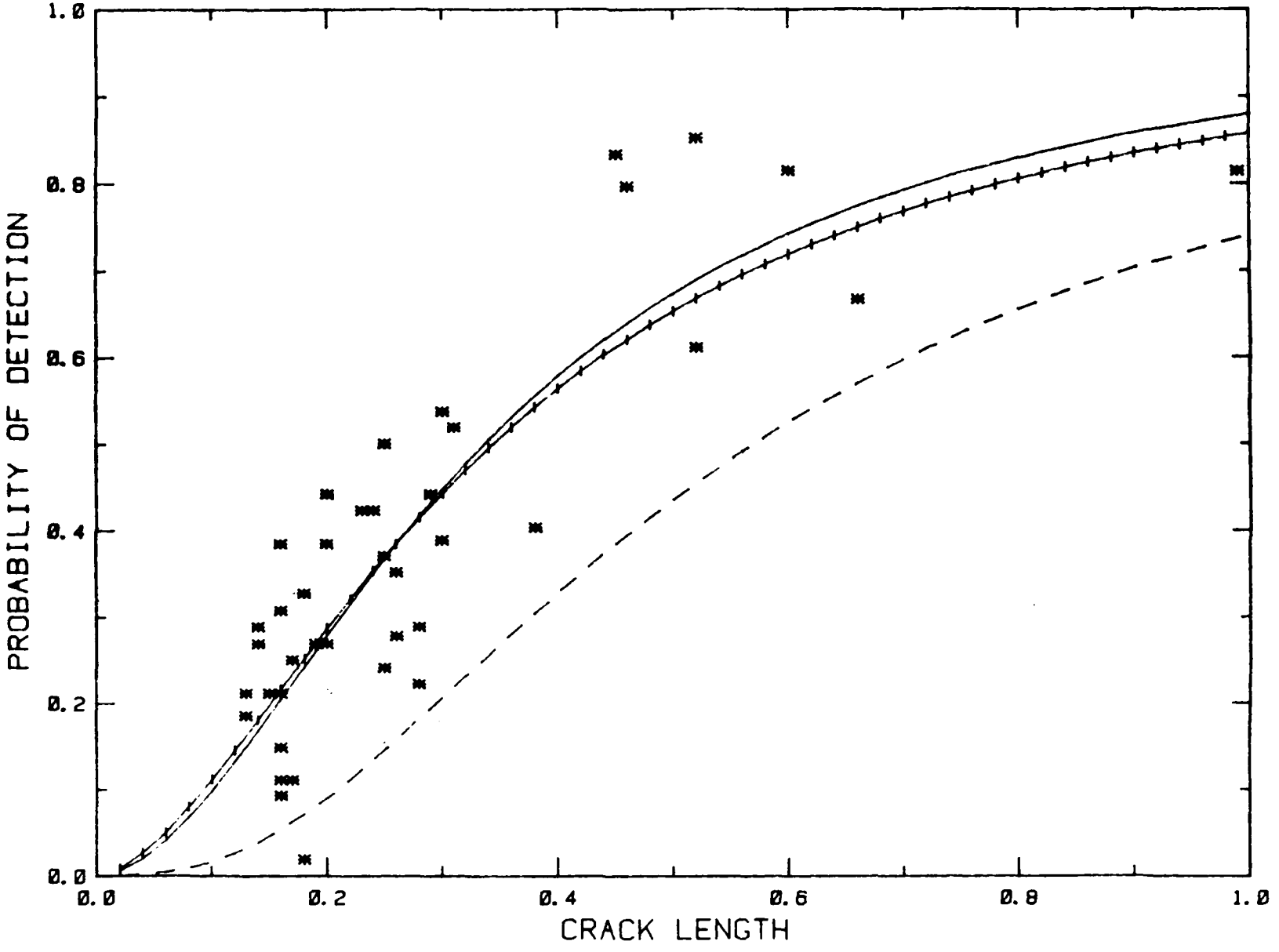


Figure 20. Comparison of Regression and Maximum Likelihood Estimates of POD for Ultrasonic Inspections of a C-130 Skin and Stringer Wing Assembly.

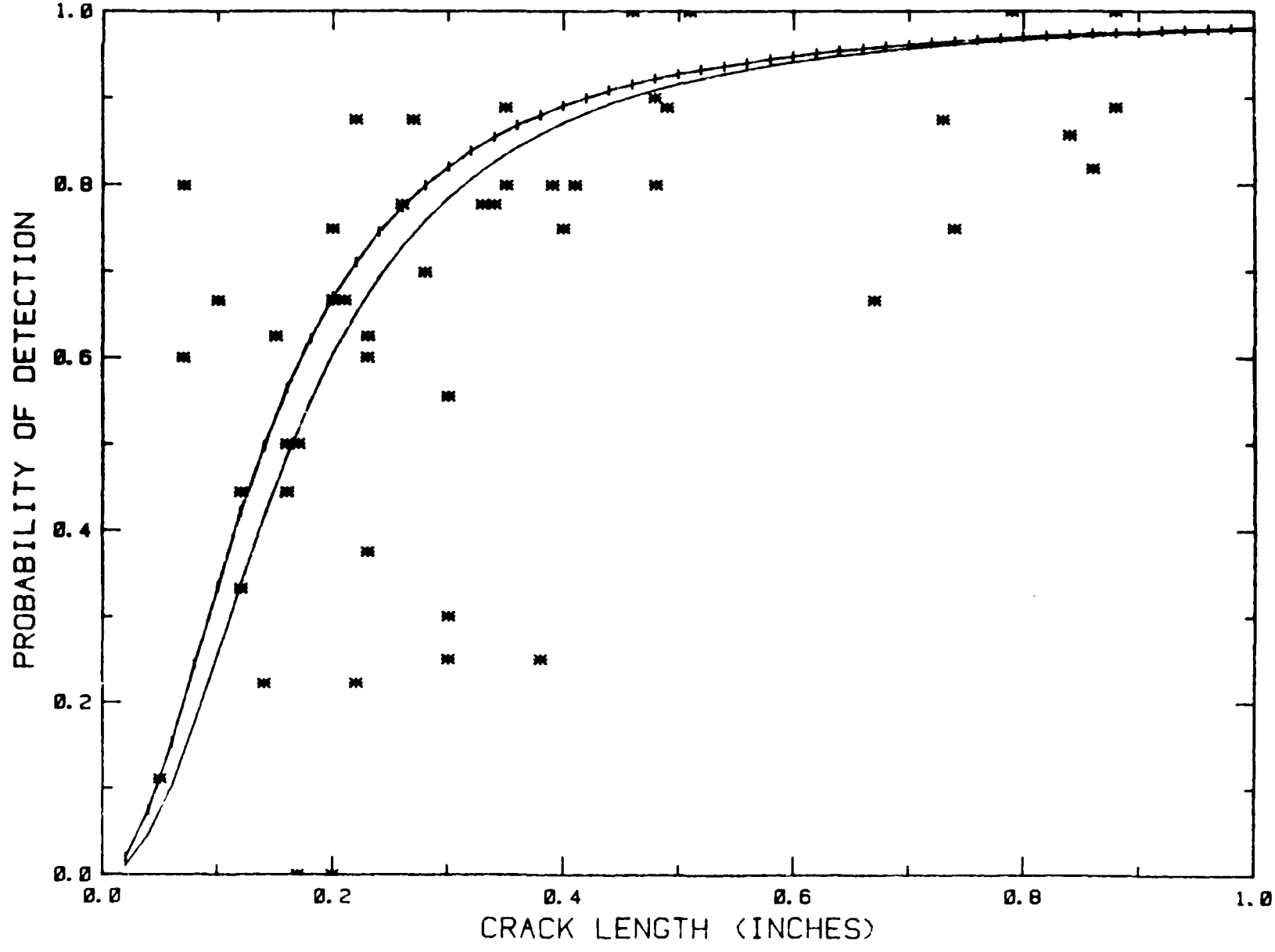


Figure 21. Comparisons of Regression and Maximum Likelihood Estimates of POD for Eddy Current Inspections of C-130 Skin & Stringer Wing Segments.

Another current problem with maximum likelihood estimates for the beta binomial model is that formulas for the confidence bounds on the POD function have not been developed and there is little motivation for developing them. Since a large number of inspections per flaw are required to get good estimates of the flaw to flaw variability in detection probabilities, regression estimates can be easily implemented. The major problem with regression estimates is that the transformation of the detection probability used to provide a linear form is undefined when the detection probability is 0 or 1. In data analysis, the transformation is performed on the observed proportion of detections for a flaw; which is commonly 0 or 1 when the number of inspections per flaw is small. Since maximum likelihood estimates also have problems for few inspections per flaw, the regression estimates are recommended.

3.3.2.2.2 Regression Analysis

The mathematical details for regression analysis of NDI reliability data are in Section A1.1 of Appendix A. The steps involved in conducting a regression analysis of NDI reliability data for the log odds model are:

1. Reduce the data to a set of pairs (a_i, p_i) where a_i is the crack length and p_i is the proportion of detections (for multiple inspections, a_i is the length of an individual crack and for grouped single inspection data a_i is the midpoint of the interval that defines the group.)

2. Transform the (a_i, p_i) pairs to (X_i, Y_i) pairs using equation (A5).

3. Calculate the standard linear regression coefficients using equations (A6) and (A7).

4. Calculate estimates of the location and scale parameters from $\hat{\alpha}$ and $\hat{\beta}$ using equation (15).

5. Calculate a lower confidence bound on the mean of Y using equation (A10).

6. Use the inverse log-odds transformation on the mean Y and the lower confidence bound on the mean Y to get estimates of the POD function and a lower confidence bound on the POD.

Figure 22 contains an example of a regression analysis applied to the same data presented in Figure 7. The points in Figure 22 represent the inspection results for a single flaw (i.e., an (a_i, p_i) pair). The mean is plotted as a solid line and the lower 95 percent confidence bound is plotted as a dotted line. Note that the lower confidence bound lies above many of the individual detection probabilities. The confidence bound is a bound on the mean of the detection probabilities not the population of detection probabilities. The POD is therefore predicted more precisely than the detection probability for an individual flaw; hence, the tight confidence bound.

3.3.3 Analysis of $\hat{a}$ Versus a Data

In the section on causes of uncertainty, it was shown that NDI uncertainty could be attributed to random variation in the response signal or $\hat{a}$ value for an NDI system. The POD can be expressed as the probability that $\hat{a}$ is bigger than the detection threshold. This section discusses the analysis of $\hat{a}$ versus a data and presents a framework for measuring the two sources of variation and for estimating the POD function.

Equation (17) supplies the basic model for the analysis of $\hat{a}$ versus a data. The flaw related and flaw independent terms c and e are random variables with means equal to 0 and variances equal to s_c^2 and s_e^2, respectively. The mean and variance of $\hat{a}$ for a single inspection of a flaw of size a picked at random are:

$$E(\hat{a}|a) = f(a), \text{ and} \tag{19}$$

$$\text{Var } (\hat{a}|a) = s_c^2 + s_e^2 \tag{20}$$

If repeated measures of a single flaw are made, the conditional mean and variance of $\hat{a}$ for the one flaw are:

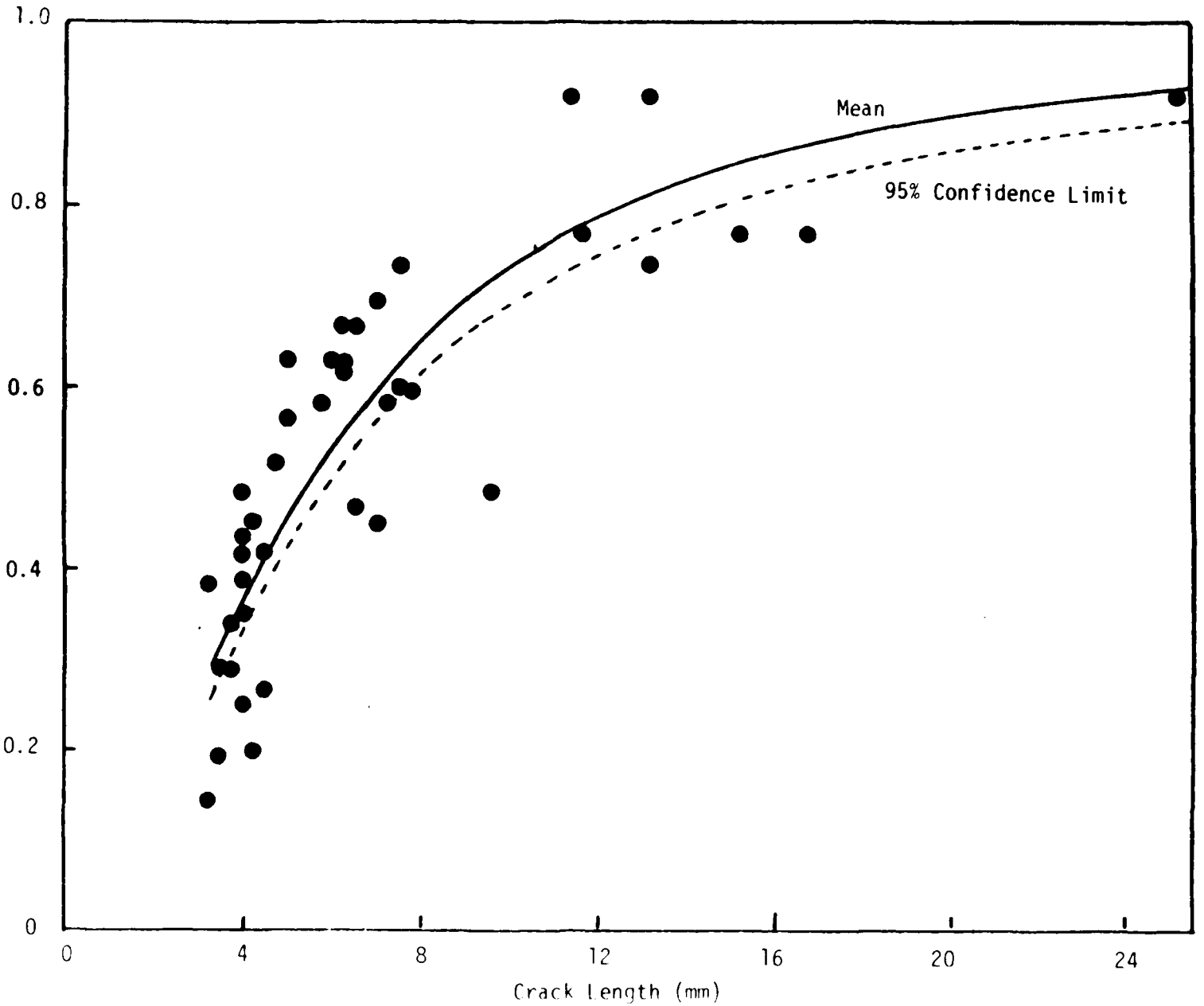


Figure 22. Example Application of Log Odds-Regression Analysis.

$$E(\hat{a}|\text{flaw}) = f(a) + c, \text{ and} \tag{21}$$

$$\text{Var}(\hat{a}|\text{flaw}) = s_e^2 \tag{22}$$

Figures 23 illustrates equations (19) and (20) while Figure 24 illustrates equations (21) and (22) with a linear function for f(a).

Figures 23 and 24 also illustrate the relationship between $\hat{a}$ and the POD function. In Figure 23 the POD as a function of a is given by the shaded area under the density function for $\hat{a}$ at a, i.e., the proportion of $\hat{a}$ values greater than the threshold for fixed a. The density functions in Figure 24 represent the distributions of $\hat{a}$ for single flaws and the shaded areas represent the detection probabilities for the individual flaws. The shaded areas in Figure 23 correspond to the POD function for the system since the POD is the mean over all flaws.

The process of transforming the distribution of $\hat{a}$ values to a POD function is further illustrated in Figure 25. The top half of Figure 25 shows the mean $\hat{a}$ curve, f(a), as a solid line and shifts of one and two standard deviations up and down as dashed lines. For a normal distribution these shifts represent the 2.3, 16, 84, and 97.7 percentiles of the $\hat{a}$ distribution. The crack lengths at which a horizontal line at a_{th} crosses the percentiles are projected down to the POD versus a axes to form the POD curve.

There are a wide variety of methods of analysing data based on the model of equation (17) with the appropriate method dependent on the form of f(a) and the number of inspections per flaw. Typically, f(a) can be converted to a linear form through convenient transformations of $\hat{a}$ and a. As an example, consider a linear relationship between $\ln(\hat{a})$ and ln(a). Equation (17) then becomes

$$Y = \alpha + \beta X + c + e, \text{ where} \tag{23}$$

$$Y = \ln(\hat{a}) \text{ and } X = \ln(a). \tag{24}$$

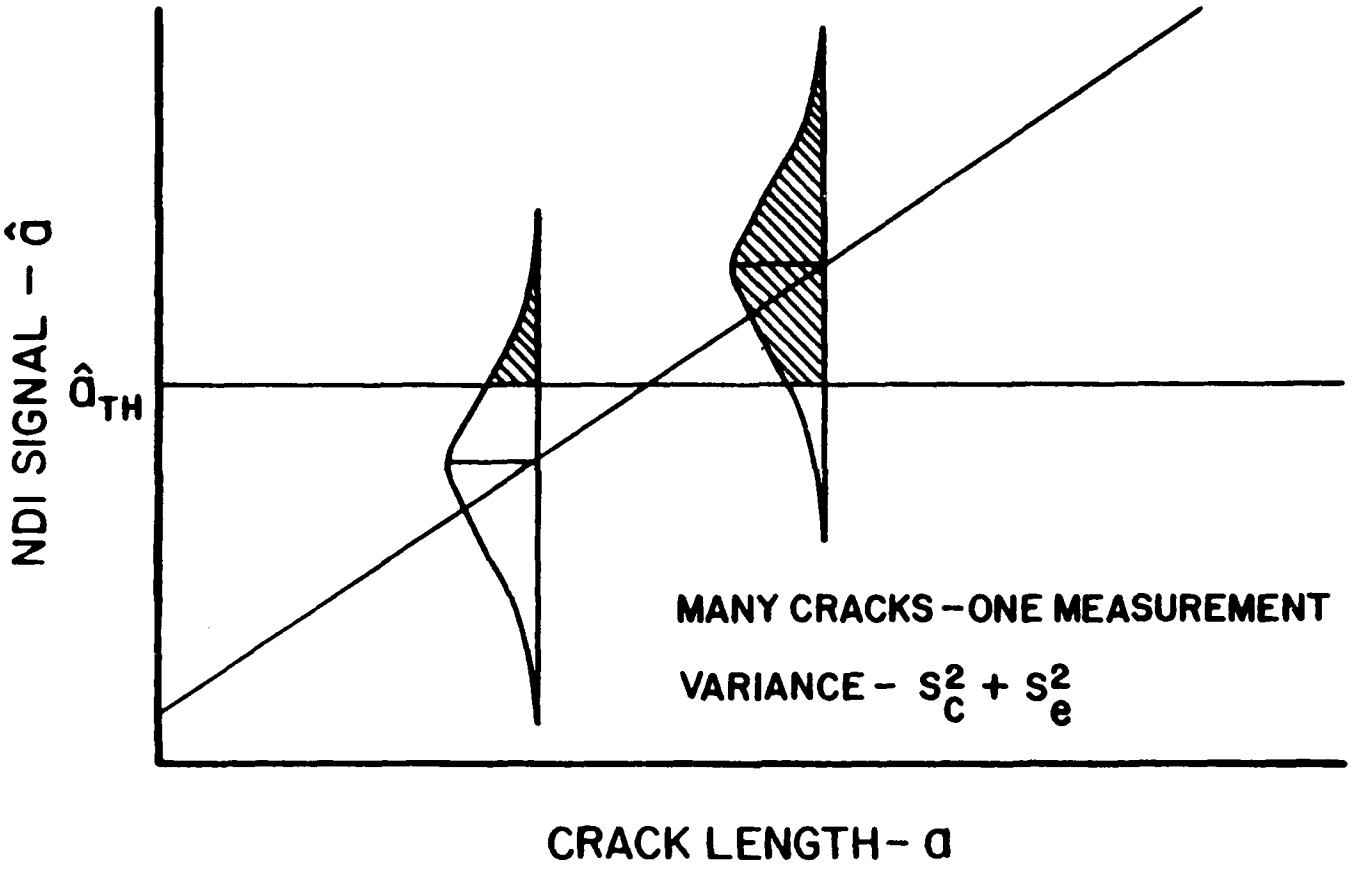


Figure 23. Example of the Distributions of Single $\hat{a}$ Values as a Function of a.

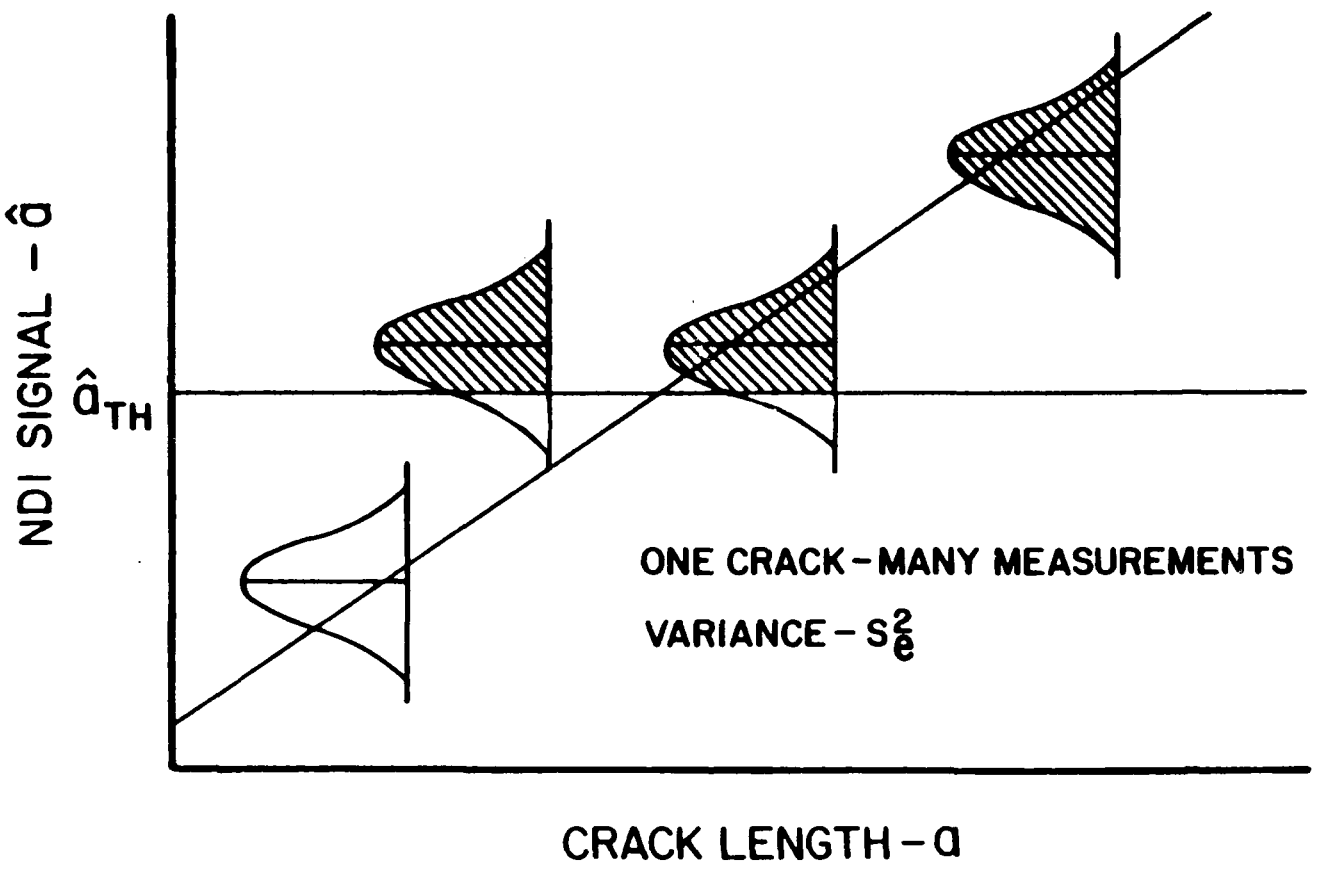


Figure 24. Example of Distributions of $\hat{a}$ for Individual Flaws.

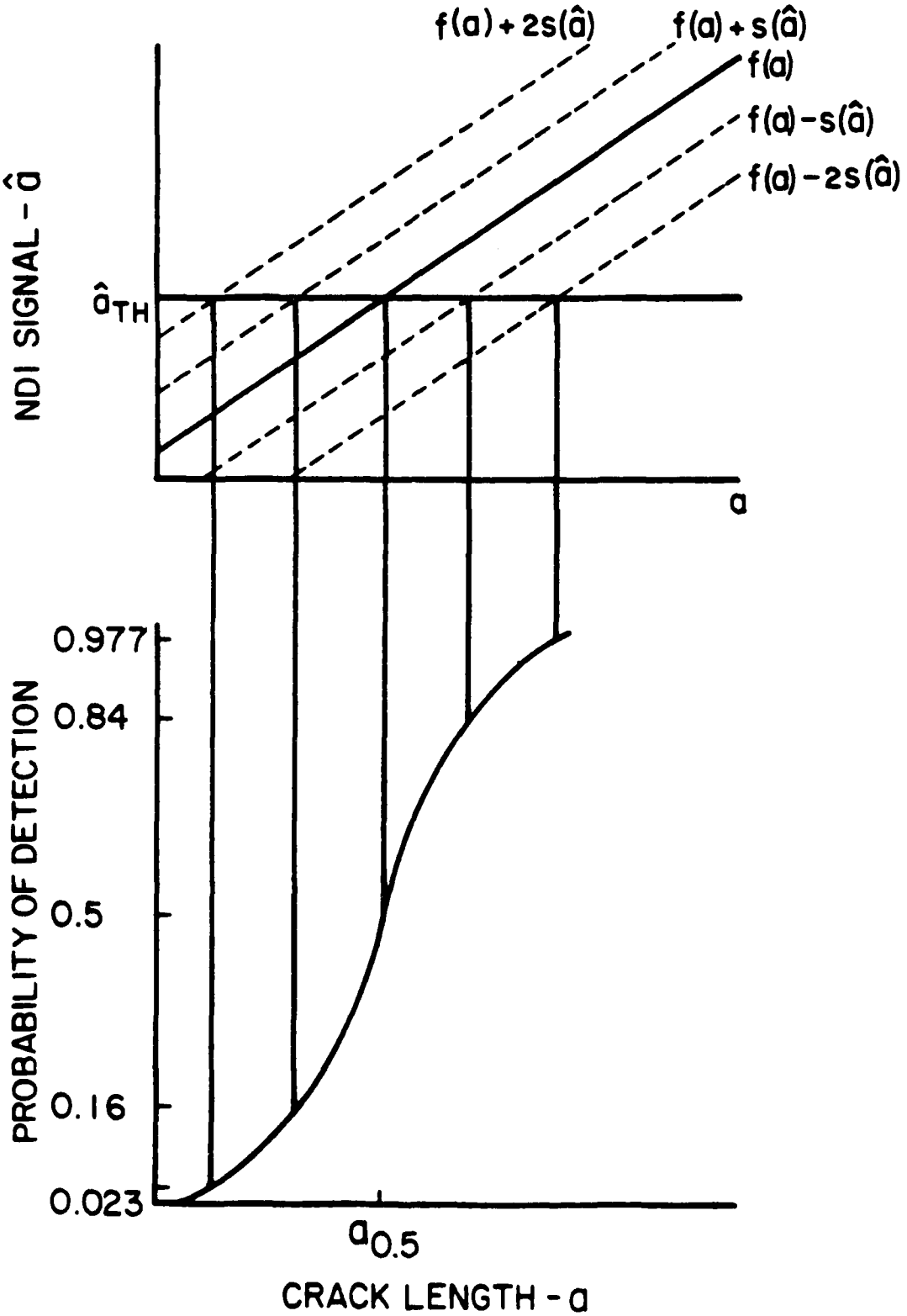


Figure 25. Projection of $\hat{a}$ Distribution to the POD Function.

If the error variables c and e are assumed to have normal distributions, the POD function is given by

$$\begin{aligned} POD(a) &= P\ (\hat{a} > a_{th}) \\ &= P\ \{\ln(\hat{a}) > \ln(a_{th})\} \\ &= P\ (Y > Y_{th}) \\ &= 1 - \Phi\ \left(\frac{Y_{th} - (\alpha + \beta X)}{s}\right) \end{aligned} \tag{25}$$

where

$$s = \sqrt{s_c^2 + s_e^2} \tag{26}$$

and $\Phi(x)$ is the standard normal distribution function. Using the symmetry properties of $\Phi(x)$, equation (25) can be reduced to

$$POD(a) = \Phi\left(\frac{X - \left(\frac{Y_{th} - \alpha}{\beta}\right)}{s/\beta}\right) \tag{27}$$

Equation (27) is the form of a lognormal distribution function with mean and standard deviation of log crack length given by:

$$\mu = (\ln(a_{th}) - \alpha)/\beta, \text{ and} \tag{28}$$

$$\sigma = s/\beta \tag{29}$$

In the previous analysis sections the log logistic function of equation (8) was used to model the POD function; however, the log logistic function is a close approximation to the lognormal distribution. The use of the lognormal distribution instead of the log logistic distribution will therefore result in very similar estimates of the POD function.

The parameters μ and σ can be estimated by substituting estimates of α, β, and s in equations (28) and (29). The parameters of the linear model, α, β, and s are estimated using the methods of Appendix A section A1.1. Equations (A6), (A7), and (A8) are used to estimate α and β and equation A10 provides the estimate for s.

The example of Equation (23) was chosen because it provided a reasonably good fit to a set of eddy current inspection data generated by the General Electric Corporation and illustrated in Figure 26. The $\hat{a}$ values in Figure 26 are percent peak voltages and the a values were normalized so that the mean $\hat{a}$ at a = 1 was 20 percent. The solid slanting line represents the mean trend in $\hat{a}$ as a function of a and the horizontal lines represent three threshold values (10, 20 and 40 percent) that were analyzed.

Equation (27) provided the estimated POD function shown in Figure 27. Increasing the detection threshold has two effects. First the median detection crack length increases with the detection threshold. Second, the slope of the POD function decreases as the detection threshold increases. The decreasing slope is the result of a constant standard deviation of log crack length, which translates to an increasing standard deviation of crack length. If the plots of Figure 27 were done on a logarithmic crack length scale, the three curves would be parallel.

A method described by Cheng and Iles (13) for calculating confidence bounds on cumulative distributions can be adapted to calculate confidence bounds on the POD given by equation (27). The formulas given by Cheng and Iles for the P percent lower confidence bound can be adapted for the estimates of μ and σ of the POD function to give:

$$POD(a) = \Phi(z_L), \text{ where} \tag{30}$$

$$z_L = \hat{z} \sqrt{\frac{\lambda}{n}\left(\frac{\hat{z}^2}{2} + \frac{(X-\bar{X})^2}{SSX} + 1\right)} \tag{31}$$

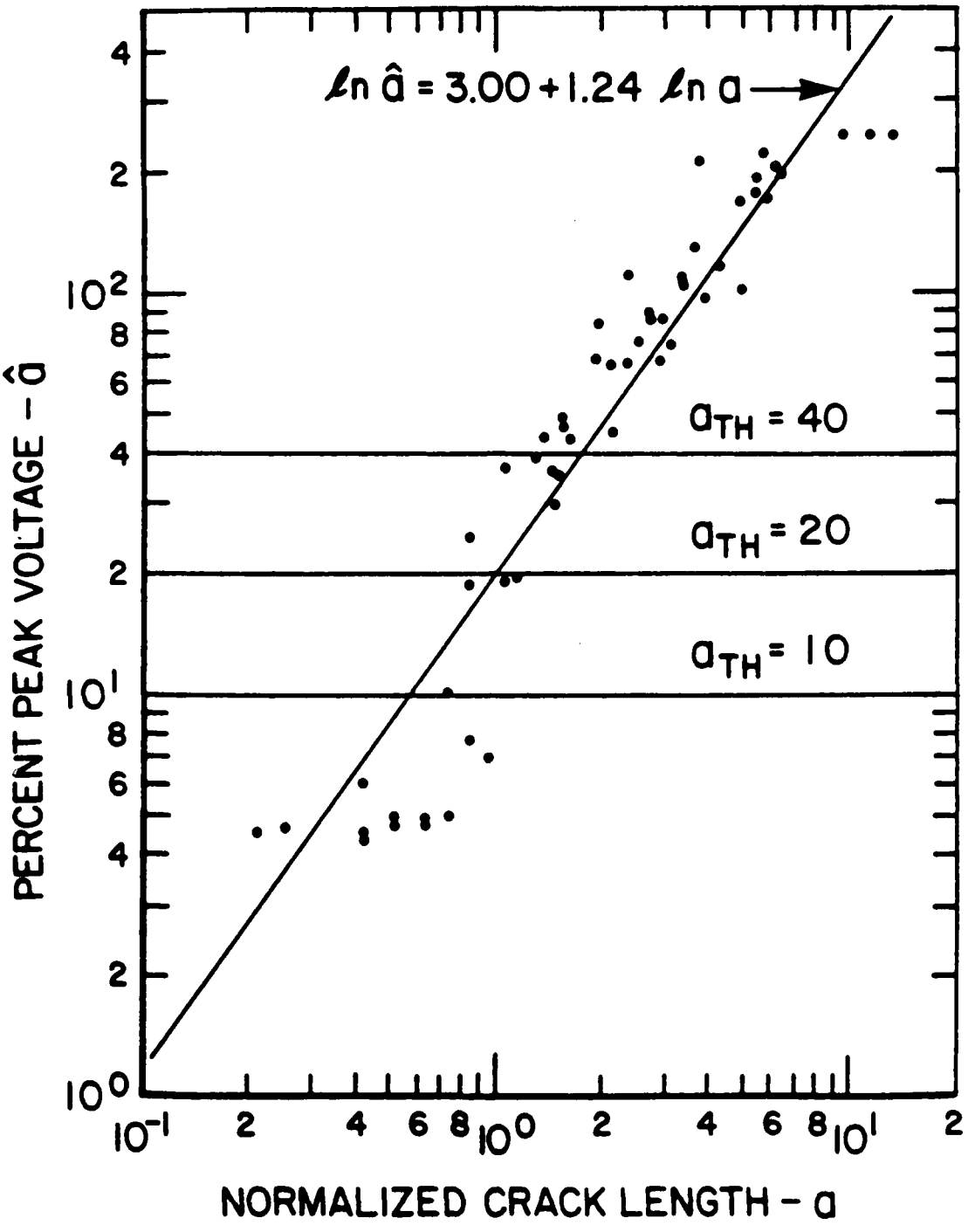


Figure 26. Example of a Versus â Eddy Current Inspection Data with a Linear Trend in ℓn â as a Function of ℓn a.

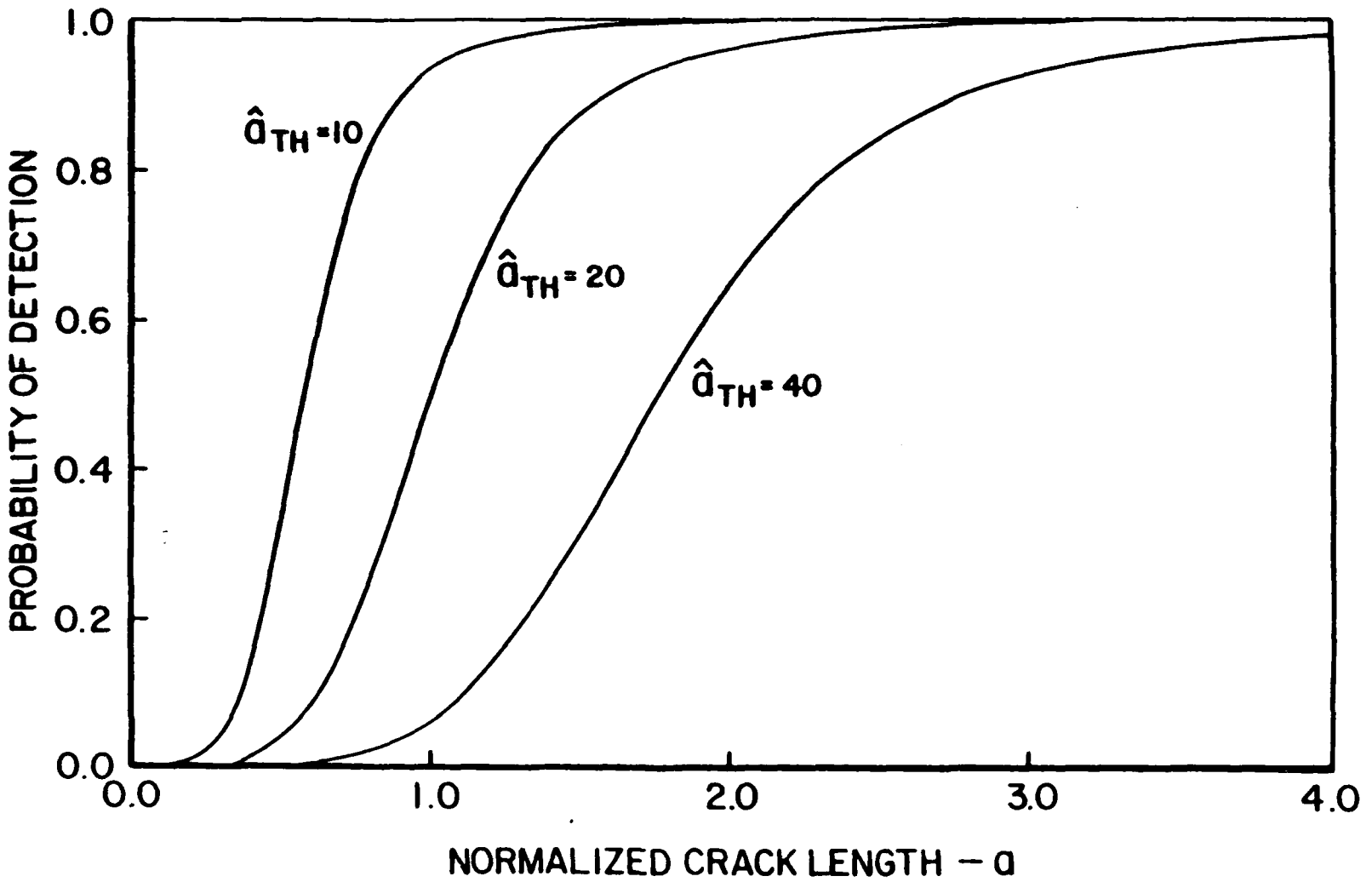


Figure 27. POD Estimates as a Function of Threshold for Data in Figure 26.

In equation (31), n is the sample size, λ is the Pth percentile of a χ^2 distribution with two degrees of freedom distribution, SSX is given by equation (A12) of Appendix A, and

$$\hat{z} = \frac{X - \hat{\mu}}{\hat{\sigma}} \tag{32}$$

An example of the 95% lower confidence bound given by equation (31) for the data of Figure 26 with a_{TH}=20 is given in Figure 28.

The analysis of the example was conducted by using only one inspection per flaw. In this case the mean square error of the regression analysis is an estimate of the combined variance $S_c^2 + S_e^2$. When more than one inspection is made of each flaw the analysis must be modified to properly estimate the individual variance components. The individual variance components, S_c^2 and S_e^2, must be estimated using a variance components model for a nested (repeated subsampling or hierarchal) experimental design. See for example, Reference 14. $\hat{S}_c^2$ and $\hat{S}_e^2$ are then added to estimate the combined variance for the POD function estimate of σ.

The problems associated with analysing NDI reliability data can be reduced by recording $\hat{a}$ values from the inspections for the analysis. There are no potentially undefined transformations and simpler models result in less complicated mathematics for estimating POD model parameters. Finally, as will be shown, the $\hat{a}$ vs a analysis is less sensitive to the distribution of crack sizes in the NDI reliability experiment.

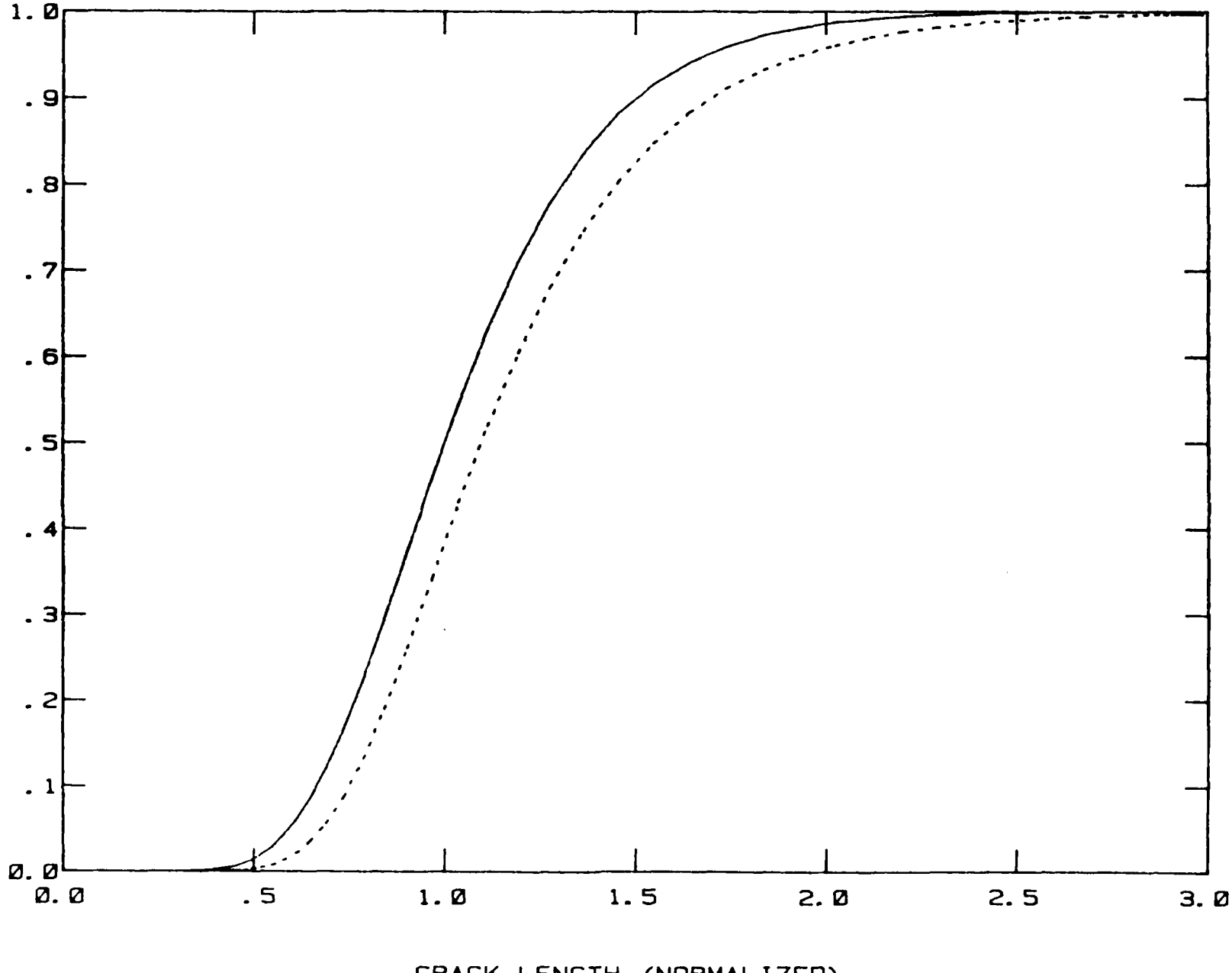


Figure 28. Mean POD with Confidence Bound Calculated from â vs a Analysis.

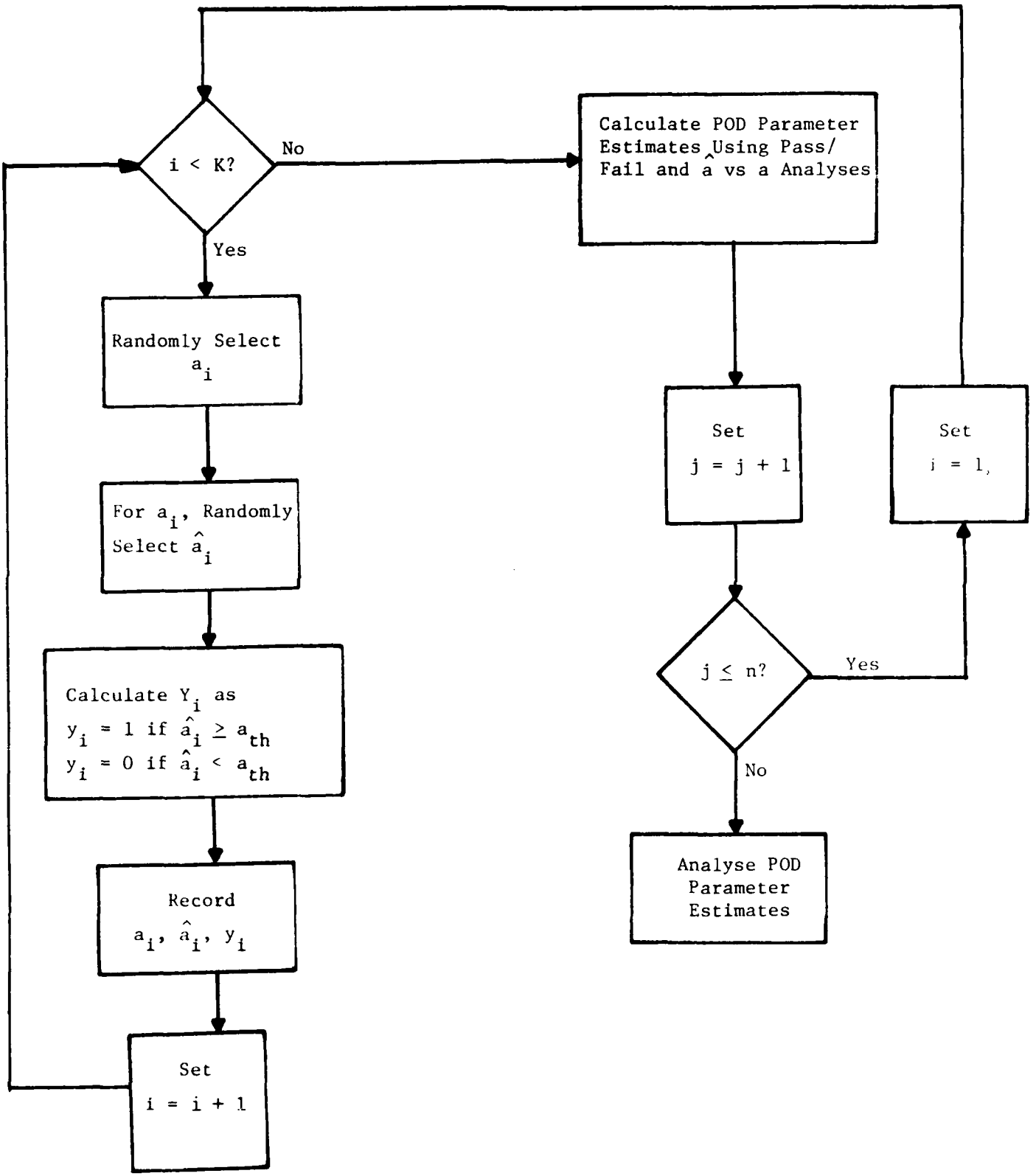
3.4 SAMPLING VARIABILITY OF POD PARAMETER ESTIMATES

This section discusses various factors in NDI reliability demonstration programs that affect the sampling variability of the estimates of the POD parameters. These factors include the type of analysis and the statistical design of the experiment. The statistical design of an experiment is the selection of the number (sample size) and sizes of the flaws to be used in the experiment. The effect of these factors on the sampling variability in POD parameter estimates was investigated through computer simulations.

The computer simulation program used in this study was modified from the program used in Reference (2). In the previous study, only the pass/fail analysis was used. A major modification required for the current study was the added simulation of $\hat{a}$ values so that the $\hat{a}$ versus a analysis could be compared with the pass/fail analysis. The modification was accomplished using equation (17) as the model of $\hat{a}$ versus a. The terms c and e were combined for the simulations since only one inspection per flaw type analyses were investigated.

A schematic diagram of the simulation program is given in Table 1. The input parameters include the POD model parameters, the demonstration program crack length distribution parameters, the sample size (n), and the number of simulations to be run (k). The program selects parameters for the $\hat{a}$ versus a model and the detection threshold to be compatable with the input POD function utilizing equation (27). A flaw size is then picked at random from a distribution of "specimen" crack lengths and the $\hat{a}$ versus a model is used to randomly generate an $\hat{a}$ value. After $\hat{a}$ has been determined, it is compared to the detection threshold to decide whether a "flaw" is present. After generating n observations, the $\hat{a}$ versus a and pass/fail analyses were performed on the simulated NDI reliability data. The results of each of the k data sets is then printed to a file so that the POD model parameter estimates can be analysed.

TABLE 1. FLOW DIAGRAM OF NDI RELIABILITY EXPERIMENT SIMULATION



In the simulated NDI reliability experiments of this study it was assumed that the crack lengths in the specimens were selected at random from a lognormal distribution with density function as given by

$$f(a) = \frac{1}{\lambda a \sqrt{2\pi}} \exp\left[-(\log a - \delta)^2 / 2\lambda^2\right] \qquad (33)$$

where

δ = mean of the natural log of crack lengths or, equivalently, the log of median crack length, and

λ = standard deviation of the natural log of crack lengths

By varying δ and λ (in comparison with μ and σ, the location and flatness parameters of the POD(a) function) the effect of the crack sizes in the NDI reliability experiment can be evaluated.

A summary of the results of the simulations are given in Table 2. The means and standard deviations of the estimates of μ and σ for both the pass/fail analysis ($\hat{\mu}_1$, $\hat{\sigma}_1$) and the $\hat{a}$ versus a analysis ($\hat{\mu}_2$, $\hat{\sigma}_2$) are given for each simulated experiment.

3.4.1 Effect of the Design of the Experiment

In this section, the design of the experiment refers to the number and distribution of the flaw sizes used in a reliability demonstration program. First, the effect of the sample size on the distribution of POD parameter estimates is discussed, then, the distribution of the flaw sizes used in the simulated NDI reliability experiments is considered, including discussions of the scatter in the flaw sizes used and the location of the center of the flaw sizes in relation to the center of the POD function.

The basic effect of sample size on the POD parameter estimates is shown in Figures 29 - 34, which contain plots of the tenth and ninetieth percentiles of the estimates of $\hat{\mu}$ and $\hat{\sigma}$ for

TABLE 2. MEANS AND STANDARD DEVIATIONS (SD) FOR POD PARAMETER ESTIMATES FROM SIMULATED NDI RELIABILITY EXPERIMENTS

Crack Lengths		POD	Sample	μ_1		σ_1		R_1	μ_2		σ_2		R_2
δ	δ	δ	Size	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
0	0.5	0.1	10	0.020	0.125	0.037	0.085	0.049	-0.006	0.036	0.096	0.028	-0.15
			30	0.012	0.051	0.071	0.064	0.056	-0.001	0.018	0.101	0.016	0.0
			100	0.000	0.027	0.100	0.027	-0.10	-0.001	0.010	0.102	0.009	-0.02
			400	0.000	0.013	0.098	0.013	0.10	0.000	0.005	0.100	0.005	-0.10
		0.25	10	0.017	0.162	0.132	0.172	-0.20	-0.004	0.096	0.243	0.093	-0.07
			30	0.001	0.088	0.214	0.099	-0.16	-0.002	0.050	0.253	0.050	0.14
			100	0.004	0.044	0.240	0.050	0.11	-0.008	0.027	0.255	0.025	-0.35
			400	-0.001	0.022	0.244	0.025	0.11	-0.002	0.012	0.248	0.014	0.004
		0.5	10	-0.033	0.523	0.457	0.566	0.11	0.014	0.355	0.610	0.691	0.54
			30	0.009	0.112	0.476	0.187	-0.05	-0.013	0.099	0.501	0.129	-0.07
			100	0.006	0.070	0.505	0.108	-0.04	-0.006	0.048	0.509	0.073	0.09
			400	-0.001	0.033	0.490	0.046	0.02	0.003	0.024	0.491	0.031	-0.06
		1.0	10	1.148	11.77	1.306	3.179	-0.02	-0.008	0.469	0.903	1.719	0.24
			30	0.159	1.092	2.471	13.66	0.90	0.003	0.282	1.186	0.684	0.26
			100	-0.010	0.135	1.025	0.344	-0.17	0.006	0.109	1.052	0.342	0.11
			400	-0.006	0.059	1.004	0.134	0.13	0.004	0.050	0.998	0.098	-0.08
		1.5	10	0.022	2.799	0.348	6.969	-0.78	0.760	7.289	3.030	27.89	0.98
			30	-0.058	0.761	2.240	2.743	-0.18	0.443	6.642	2.240	27.94	0.85
			100	-0.013	0.209	1.813	2.105	0.16	0.020	0.184	1.572	0.732	-0.25
			400	0.000	0.087	1.570	0.304	0.03	-0.001	0.079	1.526	0.248	-0.01
0	0.25	025	10	-0.002	0.198	0.179	0.709	-0.72	0.154	1.699	1.044	8.716	0.99
			30	-0.002	0.067	0.264	0.112	0.02	0.000	0.053	0.253	0.063	0.07
			100	0.006	0.036	0.244	0.053	0.06	-0.006	0.026	0.252	0.037	0.18
			400	-0.001	0.017	0.250	0.023	0.51	0.001	0.013	0.252	0.018	0.07
	1.0	1.0	10	0.057	0.656	0.400	2.301	0.41	-0.347	4.050	5.496	41.12	-0.99
			30	-0.031	0.292	1.072	0.523	-0.05	0.037	0.213	1.109	0.351	-0.34
			100	0.003	0.138	1.006	0.202	0.22	-0.003	0.116	1.012	0.136	0.02
			400	0.020	0.068	0.997	0.103	-0.01	-0.016	0.050	0.999	0.065	0.18
ln(.5)	0.5	0.5	10	-4.504	34.76	-0.096	2.866	0.11	-0.197	1.085	0.667	1.088	0.95
			30	-0.072	1.269	0.510	0.420	0.46	-0.005	0.199	0.507	0.138	-0.61
			100	0.011	0.130	0.502	0.123	0.71	-0.019	0.090	0.518	0.077	-0.71
			400	-0.004	0.059	0.498	0.062	0.68	-0.003	0.050	0.501	0.036	-0.66
ln(.5)	0.25	1	30	-6.84	64.97	-7.61	73.92	0.99	0.34	2.64	0.57	3.27	-0.95
			100	1.24	10.57	2.42	12.08	1.00	-0.16	1.21	1.20	1.73	-0.99
	0.5		30	0.31	1.35	1.68	3.05	0.92	-0.20	1.17	1.30	1.71	-0.97
			100	0.06	0.33	1.08	0.45	0.89	-0.04	0.19	1.07	0.26	-0.75
	1.0		30	0.07	0.39	1.06	0.52	0.56	-0.04	0.26	1.06	0.28	-0.38
			100	-0.004	0.15	1.02	0.24	0.38	0.02	0.12	0.99	0.14	-0.43
	1.5		30	-0.04	0.33	0.96	0.42	0.15	0.02	0.22	1.01	0.23	-0.34
			100	-0.03	0.17	0.99	0.21	0.36	0.02	0.09	1.00	0.11	-0.34
ln(.5)	0.25		30	0.80	5.60	-0.77	10.6	-0.93	0.73	9.81	2.26	15.9	0.7
			100	0.08	2.91	0.80	4.96	-0.98	0.10	0.73	1.18	1.06	0.96
	0.5		30	0.09	2.25	1.10	2.72	-0.91	0.68	0.32	1.10	0.52	0.7
			100	-0.06	0.29	1.06	0.33	-0.80	0.05	0.20	1.05	0.24	0.82
	1.0		30	-0.06	0.33	1.02	0.52	-0.47	0.04	0.21	1.02	0.26	0.36
			100	-0.03	0.17	1.05	0.23	-0.34	0.02	0.14	1.02	0.16	0.42
	1.5		30	0.02	0.36	0.97	0.47	-0.24	0.02	0.23	1.02	0.21	0.29
			100	0.01	0.14	0.98	0.19	-0.19	-0.01	0.11	0.98	0.12	0

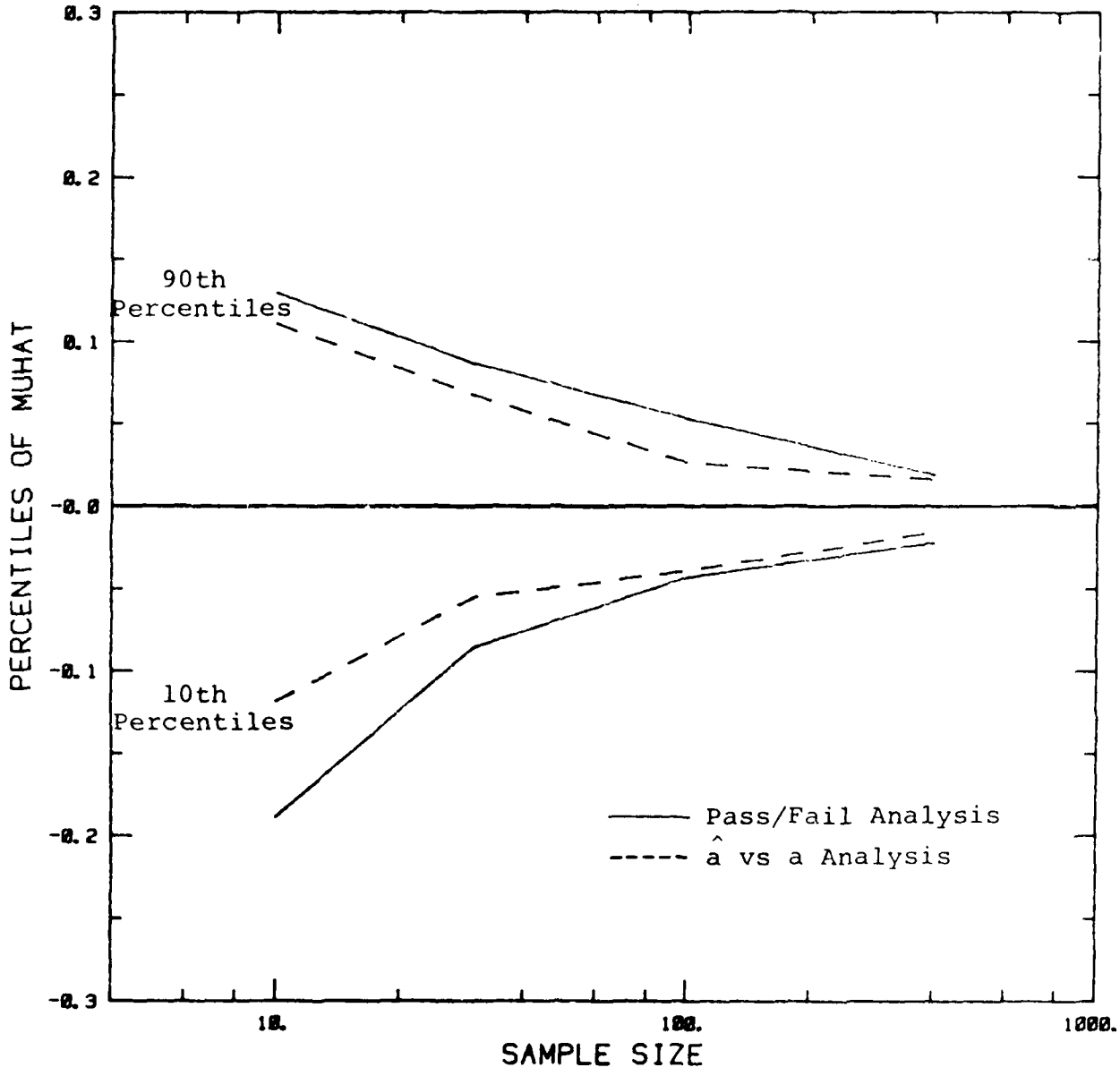


Figure 29. Tenth and 90th Percentiles of Estimates of μ as a Function of Sample Size for POD with $\mu = 0$ and $\sigma = 0.25$ Lognormal Crack Sizes with $\delta = 0$ and $\lambda = 0.25$.

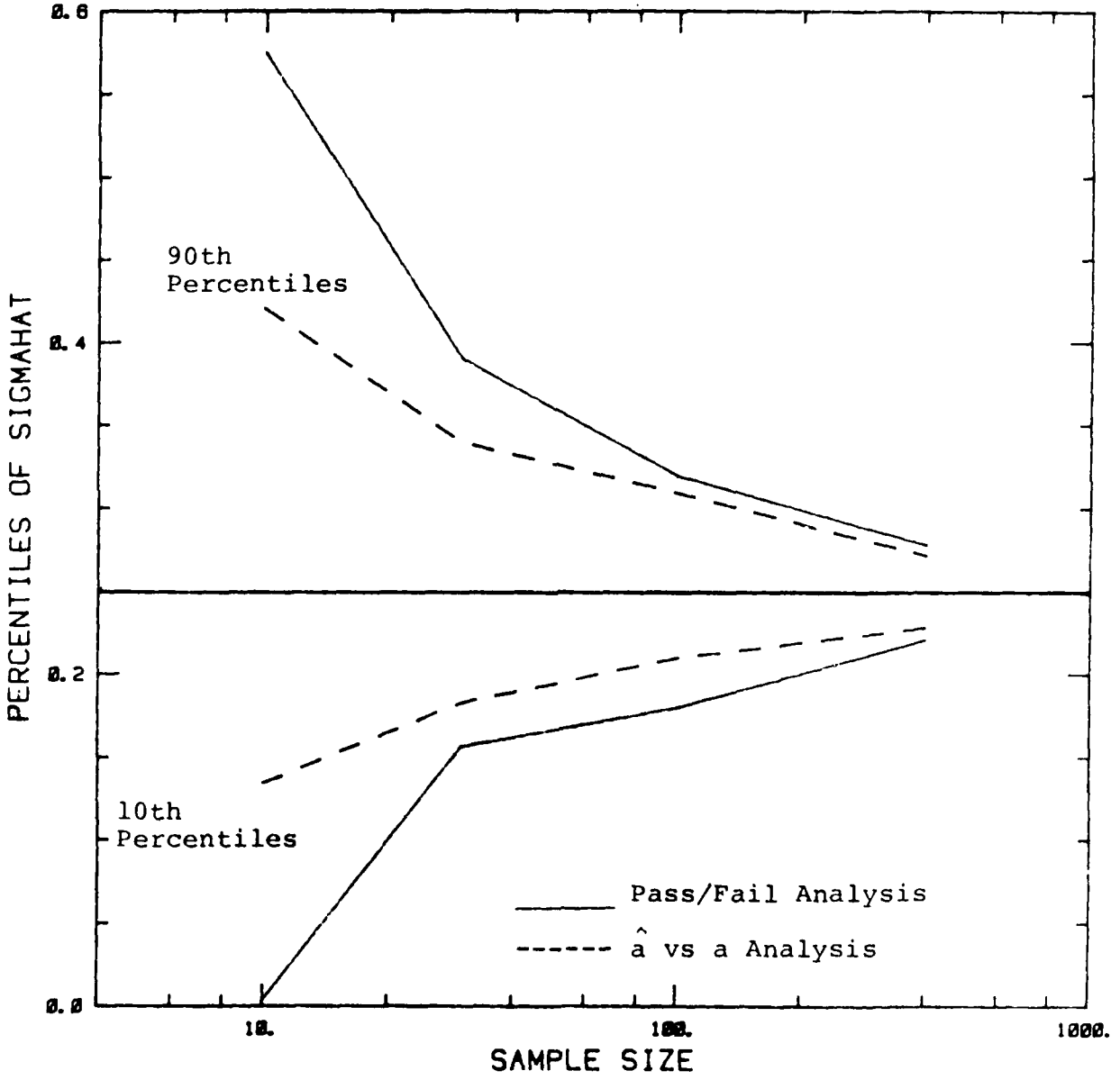
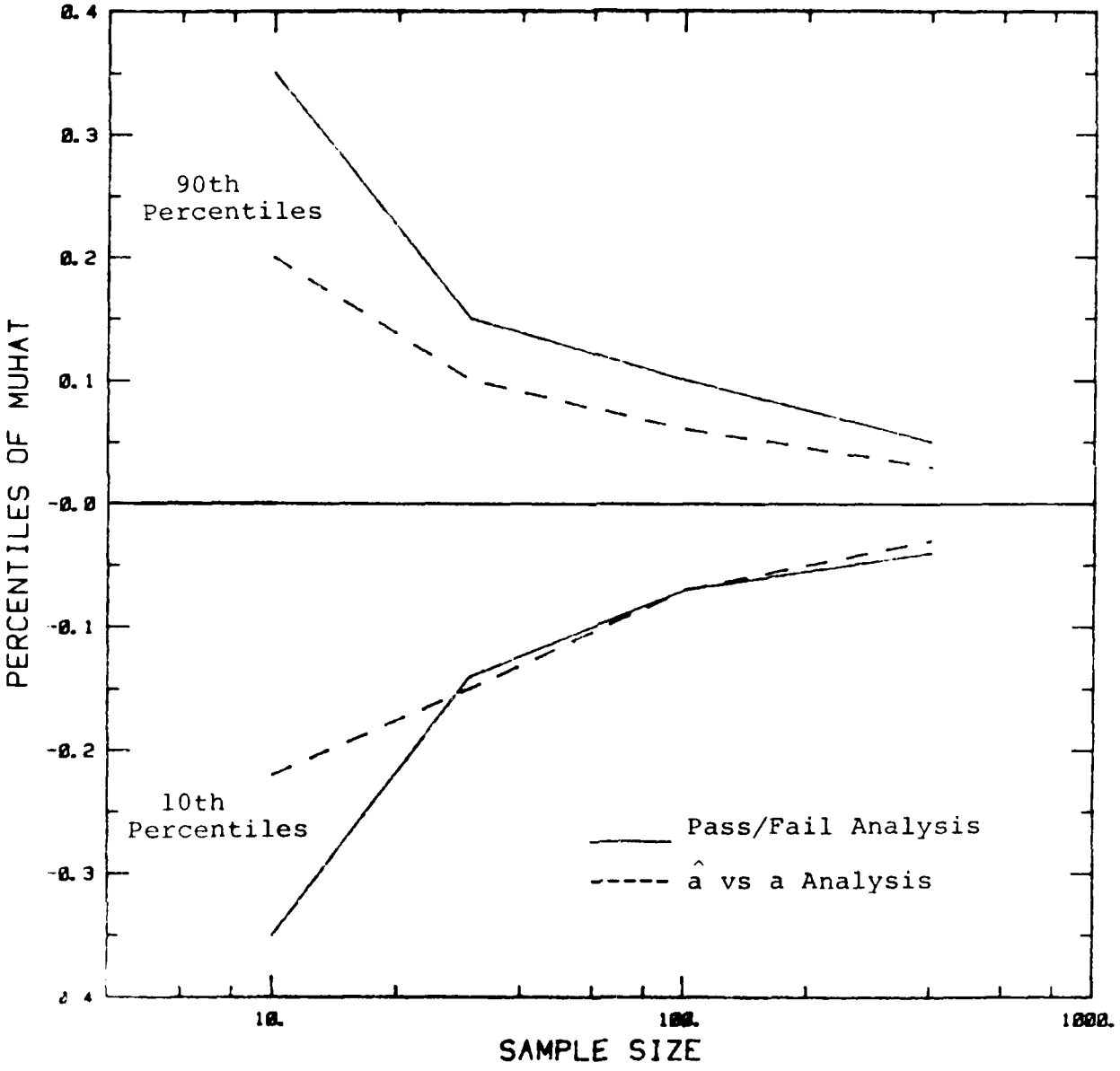


Figure 30. **Tenth and 90th Percentiles of Estimation of σ as a Function of Sample Size for POD with $\mu = 0$ and $\sigma = 0.25$ Lognormal Crack Sizes with $\delta = 0$ and $\lambda = .25$.**



31. Tenth and 90th Percentile Estimates of μ as a Function of Sample Size for POD with $\mu = 0$ and $\sigma = 0.5$ Lognormal Crack Sizes with $\delta = 0$ and $\lambda = 0.5$.

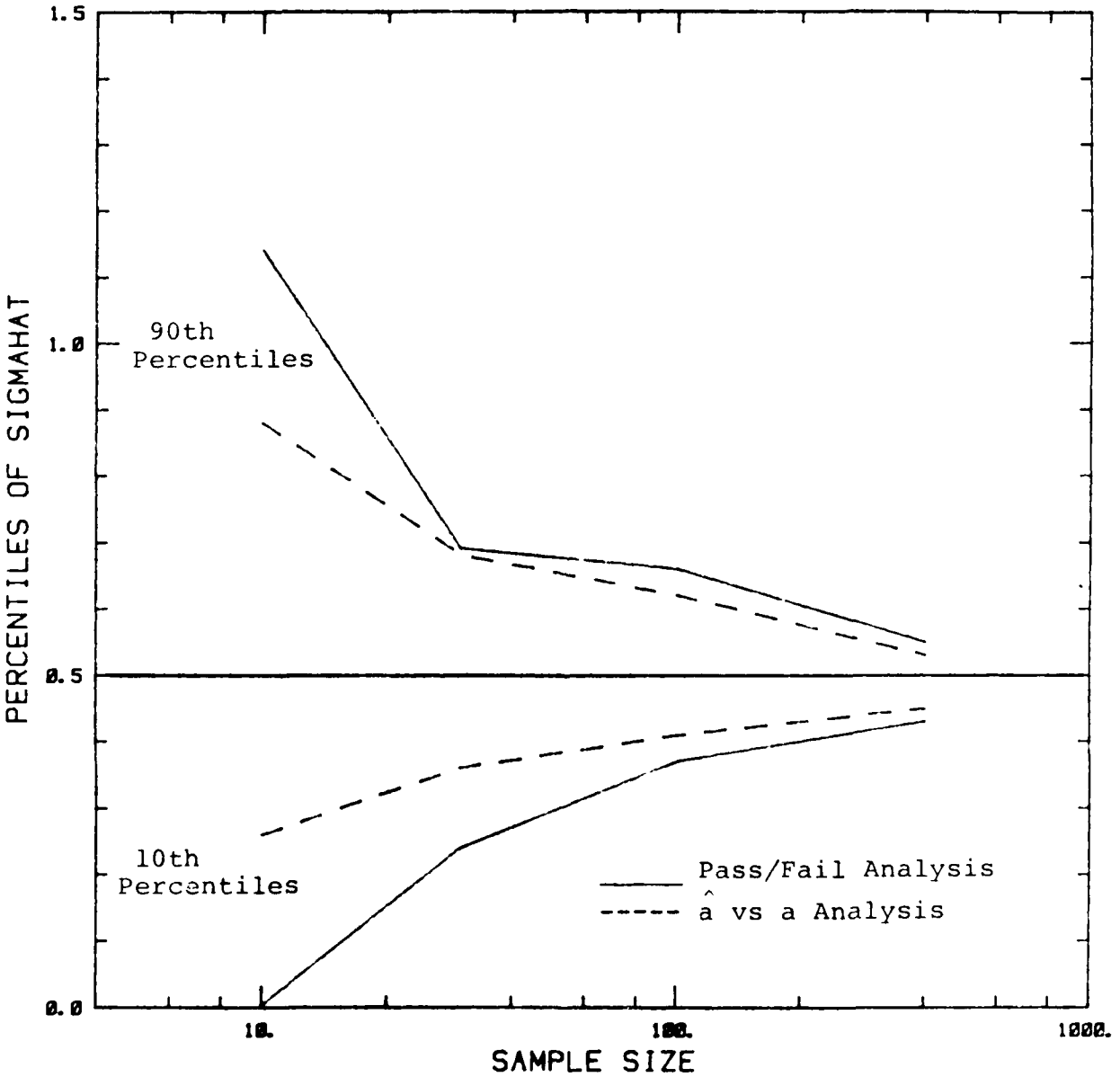


Figure 32. Tenth and 90th Percentile Estimates of σ as a Function of Sample Size for POD with $\mu = 0$ and $\sigma = 0.5$ Lognormal Crack Sizes with $\delta = 0$ and $\lambda = 0.5$.

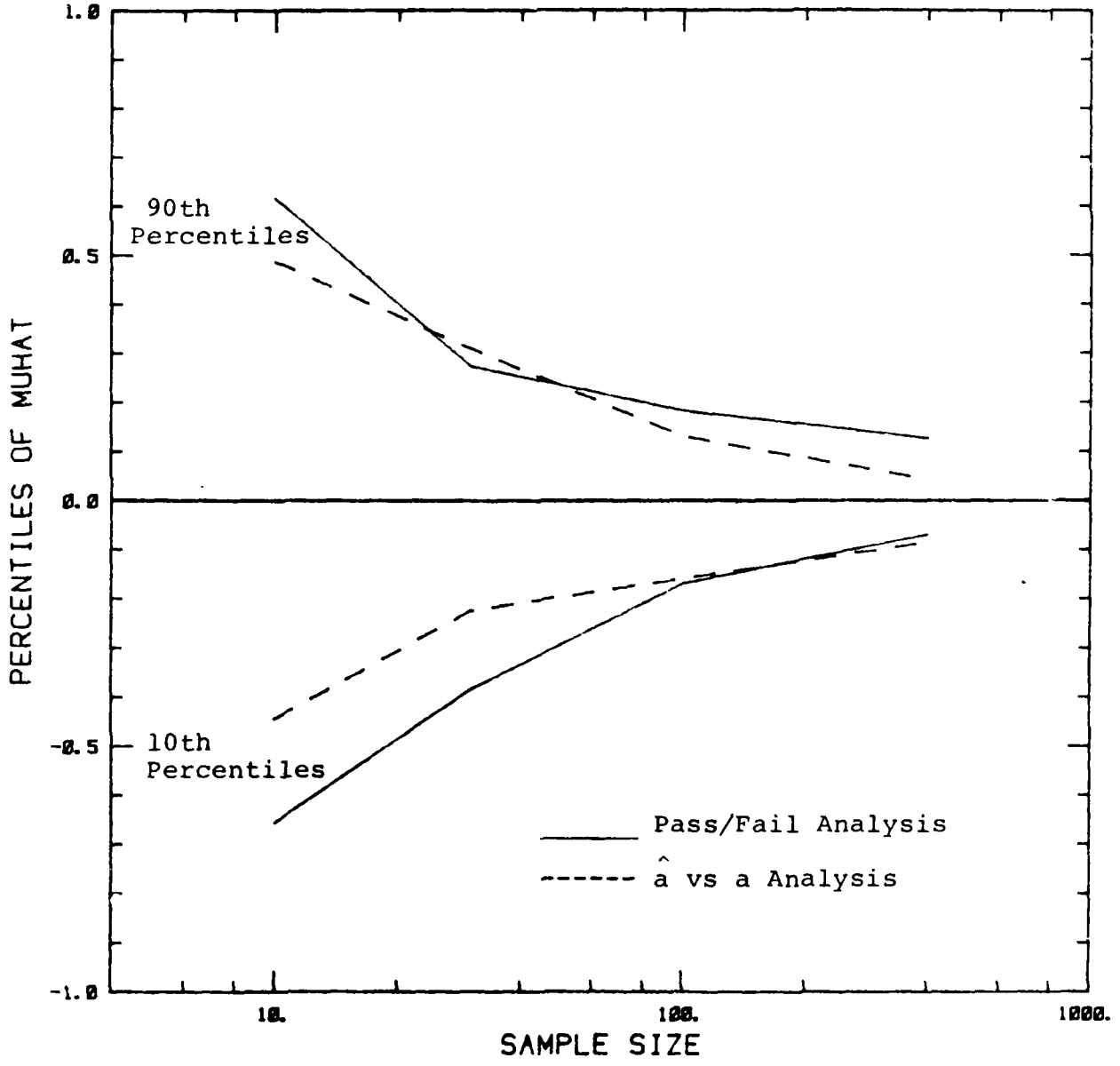


Figure 33. Tenth and 90th Percentile Estimates of μ as a Function of Sample Size for POD with $\mu = 0$ and $\sigma = 1$ Lognormal Crack Sizes with $\delta = 0$ and $\lambda = 1.0$.

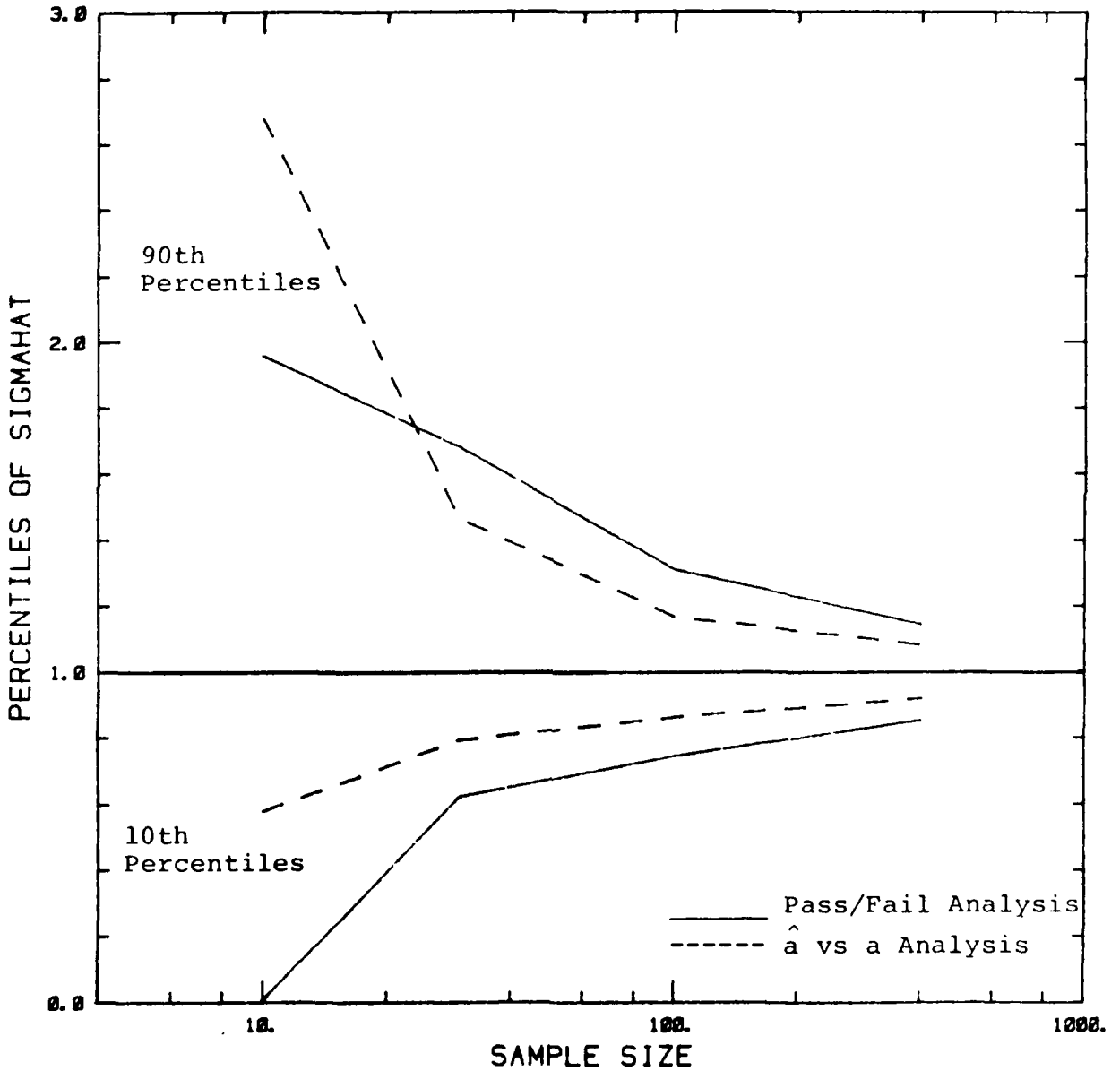


Figure 34. Tenth and 90th Percentile Estimates of σ as a Function of Sample Size for POD with $\mu = 0$ and $\sigma = 1$ Lognormal Crack Sizes with $\delta = 0$ and $\lambda = 1.0$.

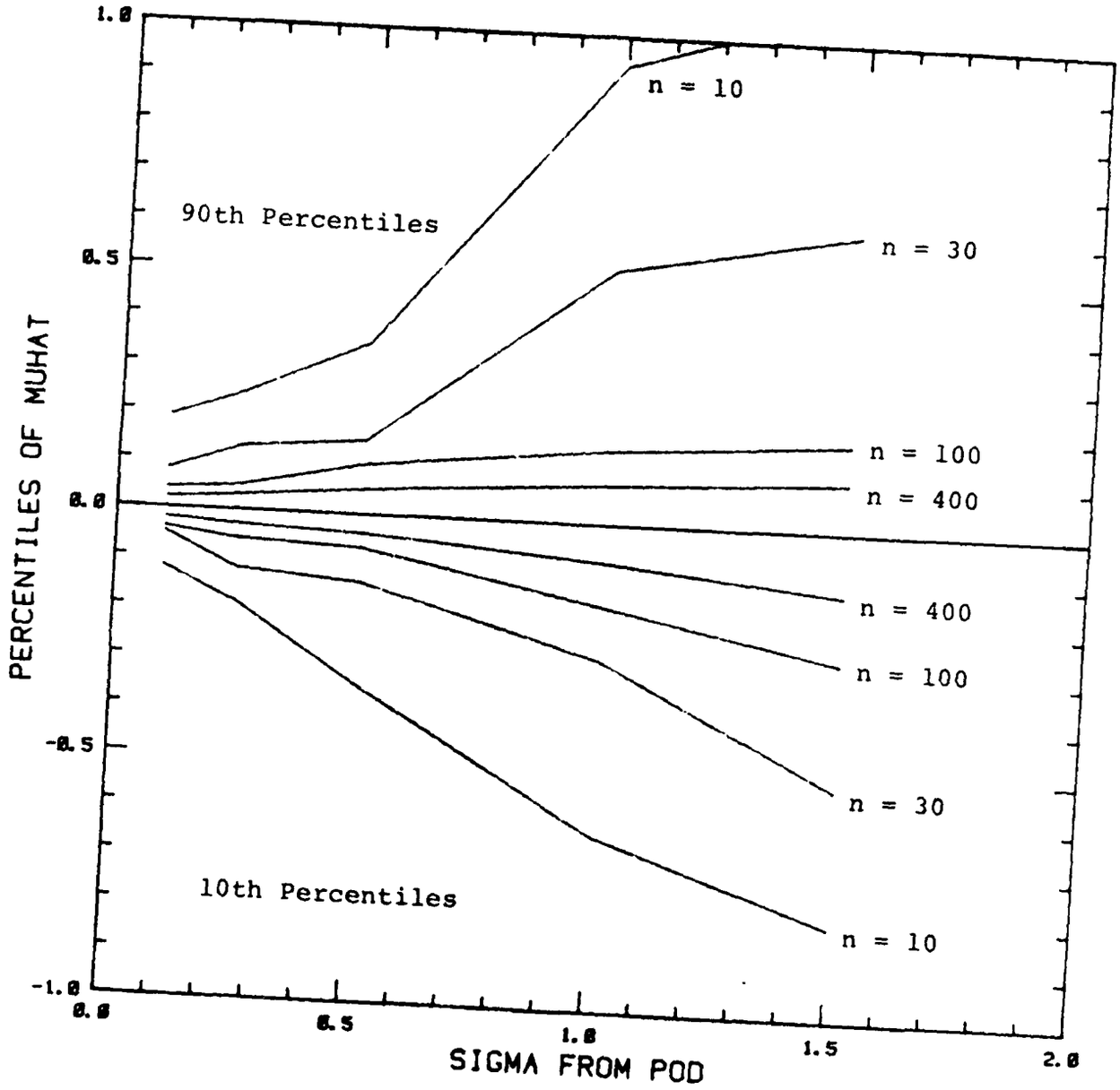


Figure 35. Tenth and 90th Percentiles of Estimates of μ as a Function of σ for Selected Sample Sizes - Pass/Fail Analysis.

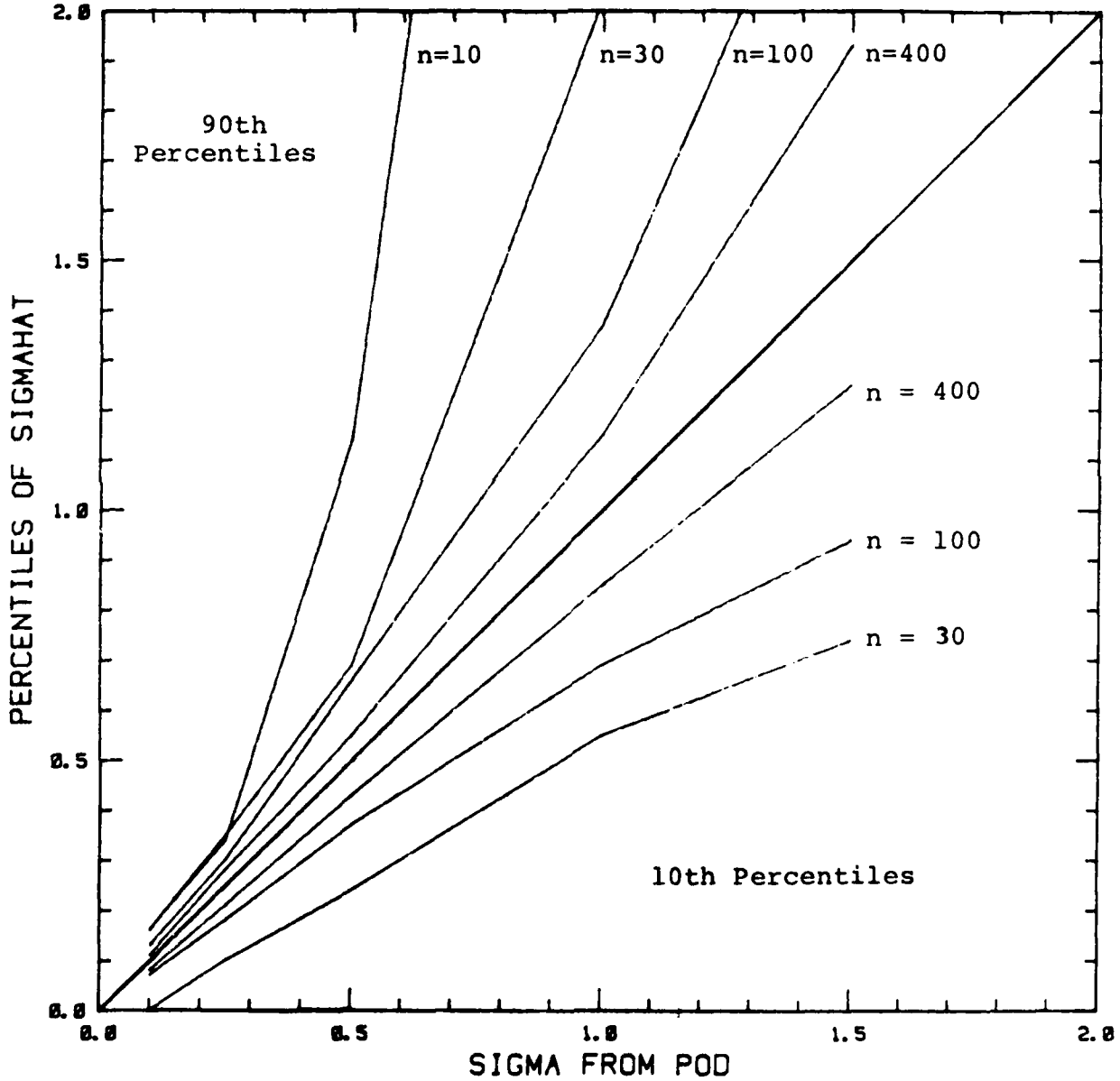


Figure 36. Tenth and 90th Percentiles of Estimates of σ as a Function of σ for Selected Sample Sizes - Pass/Fail Analysis.

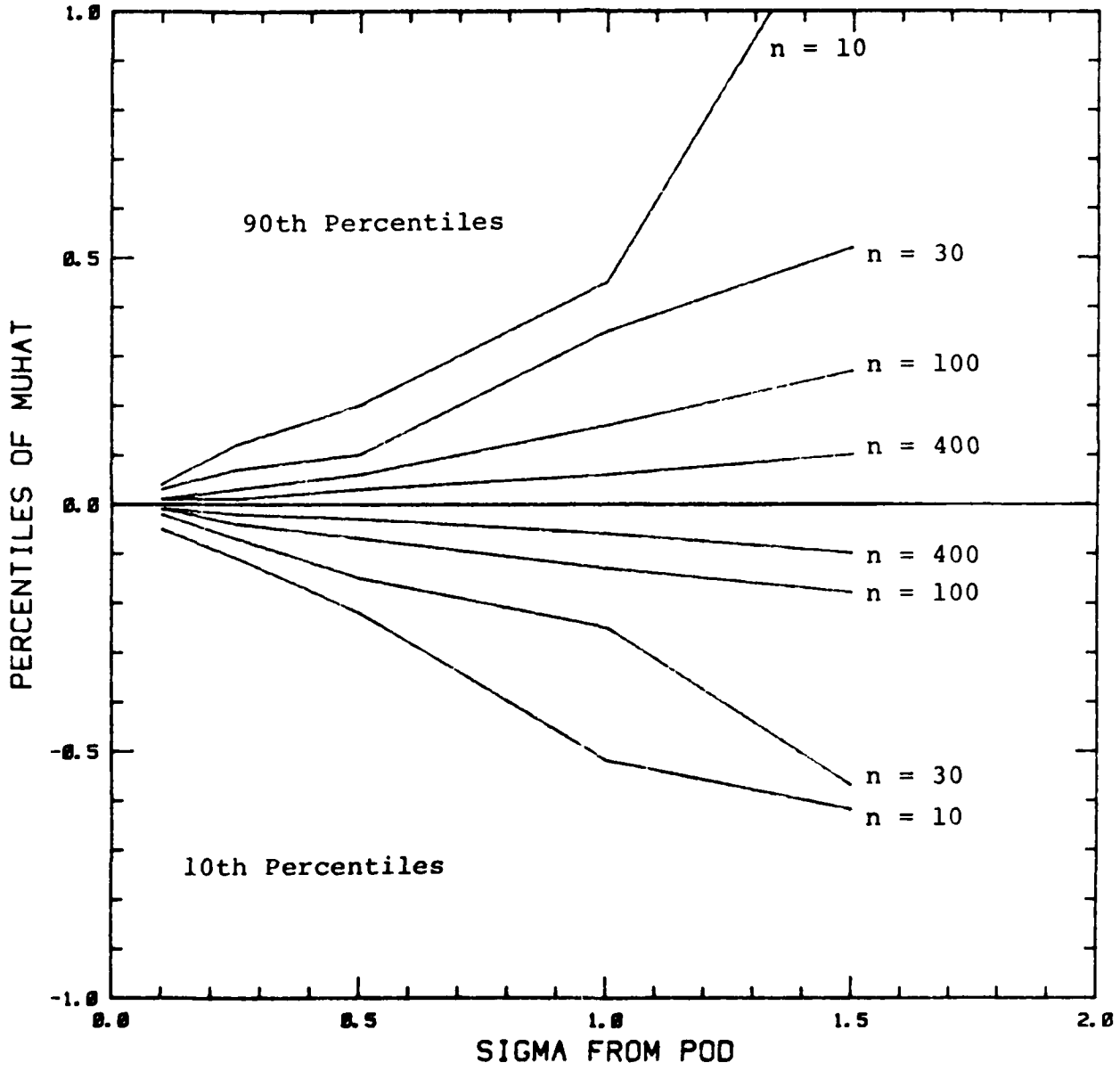


Figure 37. Tenth and 90th Percentiles of Estimates of μ as a Function of σ for Selected Sample Sizes - $\hat{a}$ vs a Analysis.

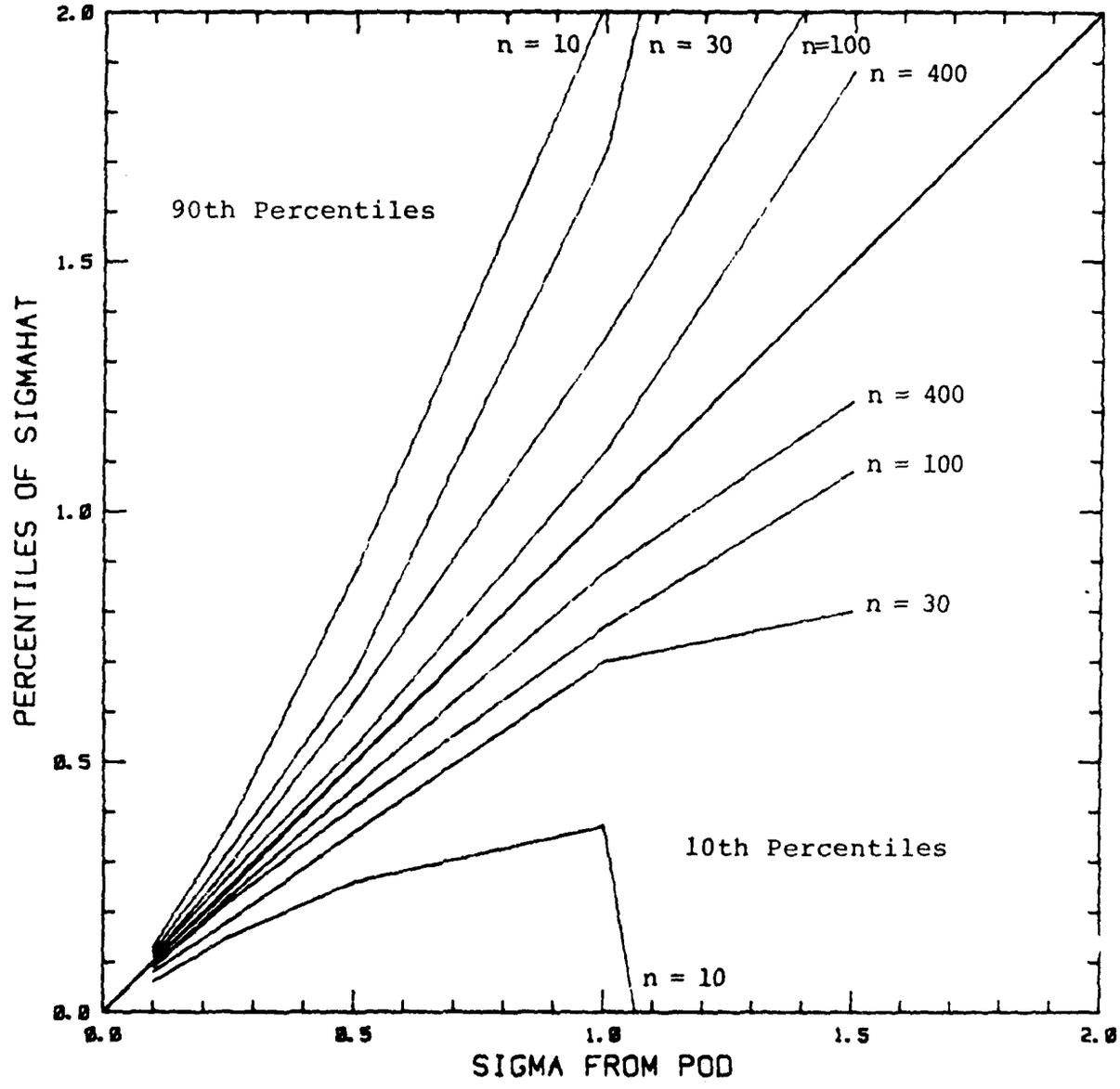


Figure 38. Tenth and 90th Percentiles of Estimates of σ as a Function of σ for Selected Sample Sizes - $\hat{a}$ vs a Analysis.

both the pass/fail analysis and the $\hat{a}$ versus a analysis as a function of sample size. The range of parameter estimates and therefore the scatter, decreases with sample size. The standard deviations of $\hat{\mu}$ and $\hat{\theta}$ are theoretically proportional to $1/\sqrt{n}$ and the simulation results generally agree with this reduction.

Another view of the sample size effect as a function of the flatness of the POD as measured by σ is shown in Figures 35-38. The curves in Figures 35-38 are plots of the tenth and ninetieth percentiles for the sample sizes indicated. The pass/fail and $\hat{a}$ versus a analyses are presented in separate figures. The scatter in the POD parameter estimates increases with σ; that is, there is less scatter in the estimates for steep POD functions than for flat POD functions.

A steep POD function gives more discriminatory power than a flat POD function in the sense that a small change in crack length will result in a larger change in POD for the steep POD than for the flat POD. Since small σ values result in steep POD functions, the smaller variance of parameter estimates for small σ is a result of the better discriminatory power. The steepness of the POD function is not, however, a design consideration since σ is one of the parameters to be estimated.

Figures 39 and 40 also show the effect of sample size on the distribution of POD parameter estimates with a different design. In Figures 39 and 40 , the center of the flaw size distribution is 0.5 while the center of the POD function is 1. The shift in the flaw sizes results in an increase in the scatter of the POD parameter estimates. The cause of the increased variability in the estimates is that for small sample sizes, the range of flaw sizes in a sample may not adequately span the POD function. Because of the shift in location of the flaw sizes, there will be fewer flaws in the center of the POD function.

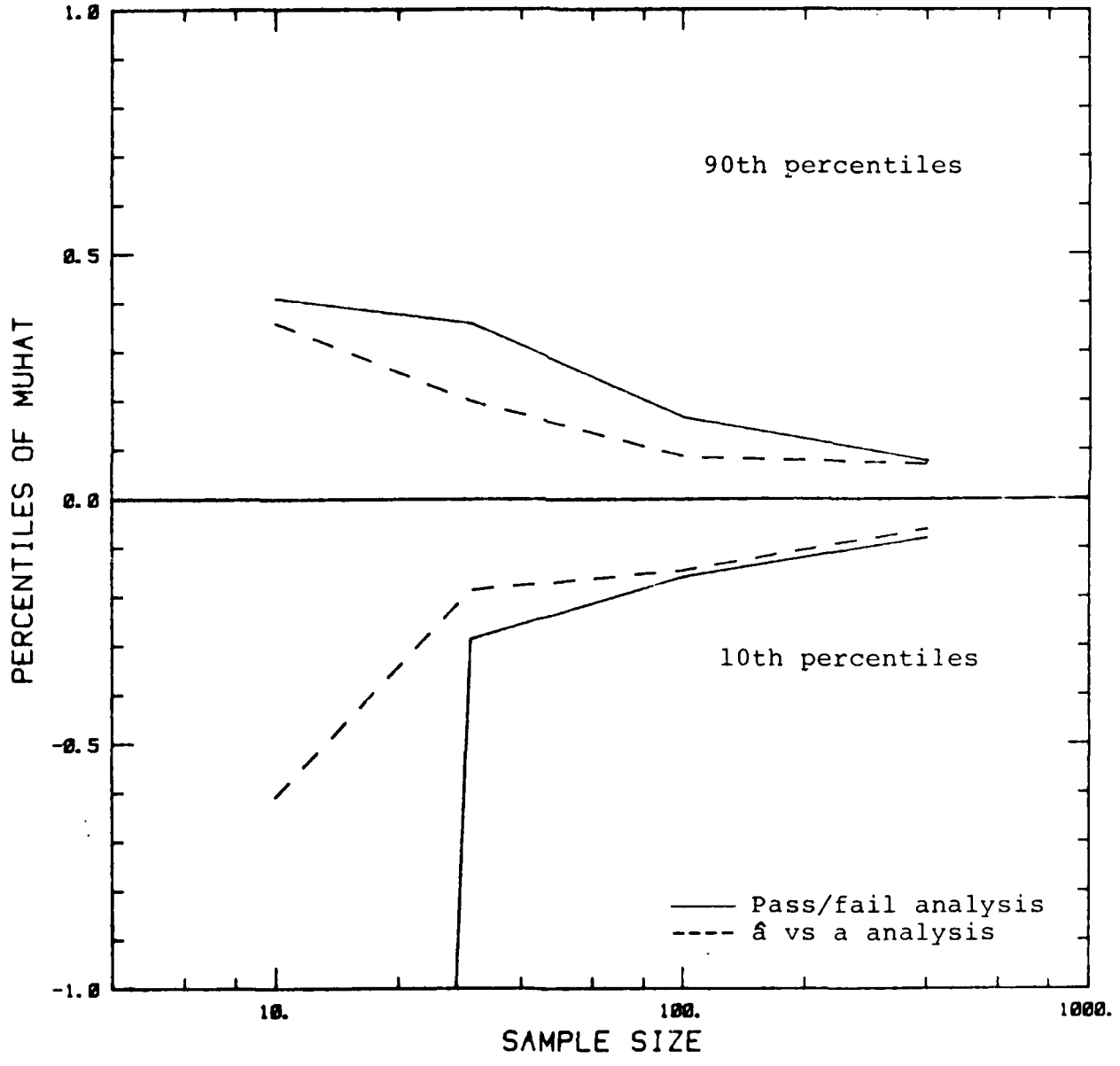


Figure 39. **Tenth and 90th Percentiles of Estimates of μ as a Function of Sample Size for POD with $\mu = 0$ and $\sigma = 0.5$ - Lognormal Crack Sizes With $\delta = \ln 0.5$ and $\lambda = 0.5$.**

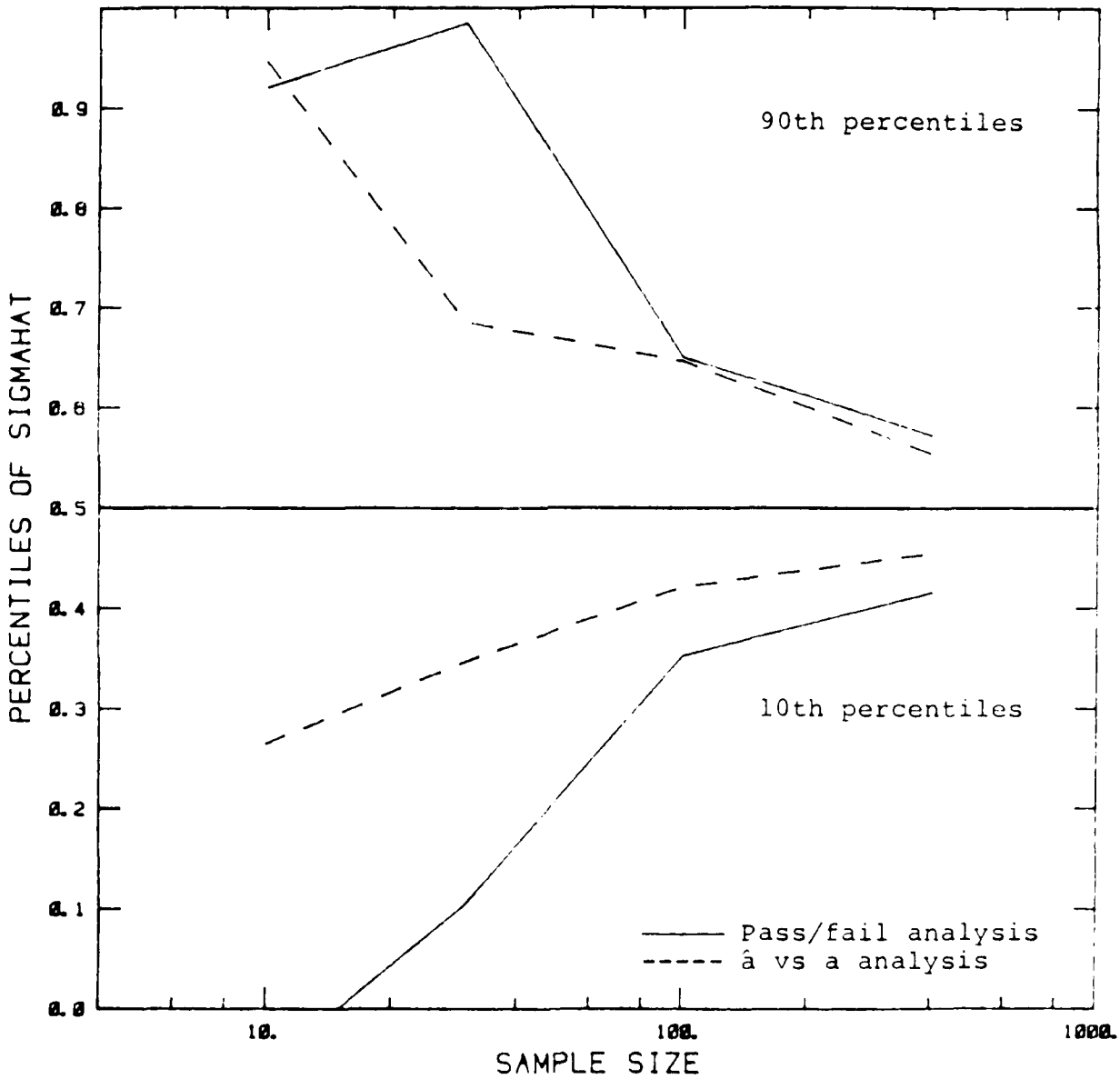


Figure 40. Tenth and 90th Percentiles of Estimates of σ as a Function of Sample Size for POD with μ = 0 and σ = 0.5 - Lognormal Crack Sizes with δ = ln 0.5 and λ = 0.5.

Figures 41 - 48 contain plots of percentiles of $\hat{\mu}$ and $\hat{\sigma}$ as a function of the standard deviation of the flaw sizes used in the simulation. In each case there is a large amount of scatter for small standard deviations. As the standard deviation increases, the range of flaws used in an experiment increases. After the range of flaw sizes increases sufficiently to span the POD function, the scatter in the parameter estimates settles down to a fixed level.

The foremost consideration in designing an NDI reliabilty demonstration program should include flaw sizes that span the full range of POD values from 0 to 1. The distributions of parameter estimates from the simulations were seen to contain a large number of 'outliers' when the flaws fell in one tail of the POD function. Reasonable distributions for $\hat{\mu}$ and $\hat{\sigma}$ were obtained when the distribution of flaw sizes is similar to the POD function. Fairly stable distributions for $\hat{\mu}$ and $\hat{\sigma}$ were obtained when flaw sizes spanned the POD function and the sample size was 30 or larger.

3.4.2 Type of Analysis

The $\hat{a}$ versus a analysis is compared to the pass/fail analysis on the basis of accuracy and precision, two measures of an analysis method's ability to estimate a parameter. In estimation, accuracy refers to the difference between the expected value, or mean, of the estimate of the parameter and the true value. Often, the expected value of a parameter estimate is the parameter itself, expressed as

$$E(\hat{\theta}) = \theta \tag{34}$$

If the expected value of $\hat{\theta}$ is not θ, the difference between the two is called the bias and is given by

$$b = E(\hat{\theta}) - \theta \tag{35}$$

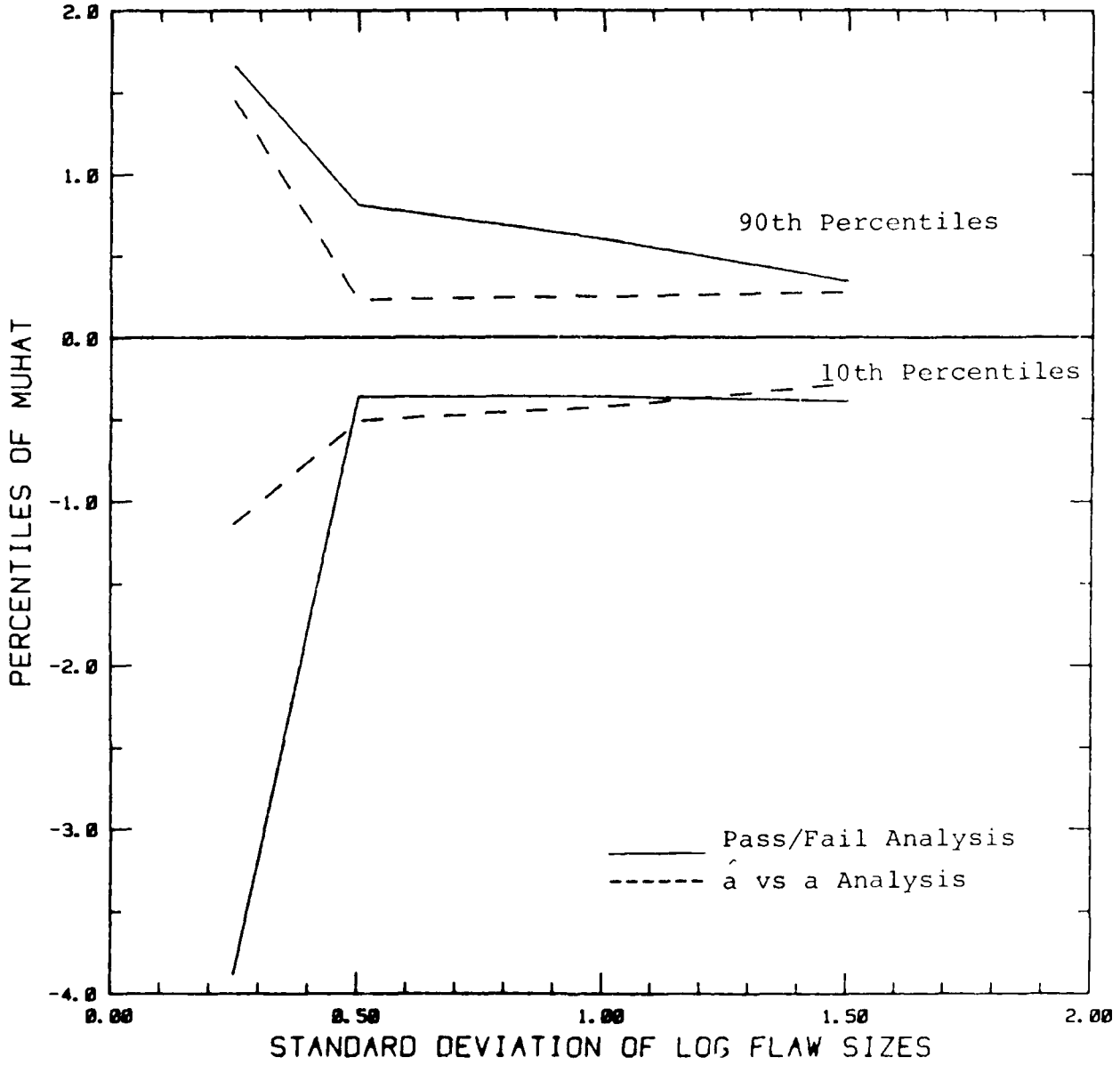


Figure 41. Tenth and 90th Percentiles of Estimates of μ as a Function of Standard Deviation of Log Flaw Sizes With Average Log Flaw Size of 0.5 and Sample Size of 30 - POD σ= 1.

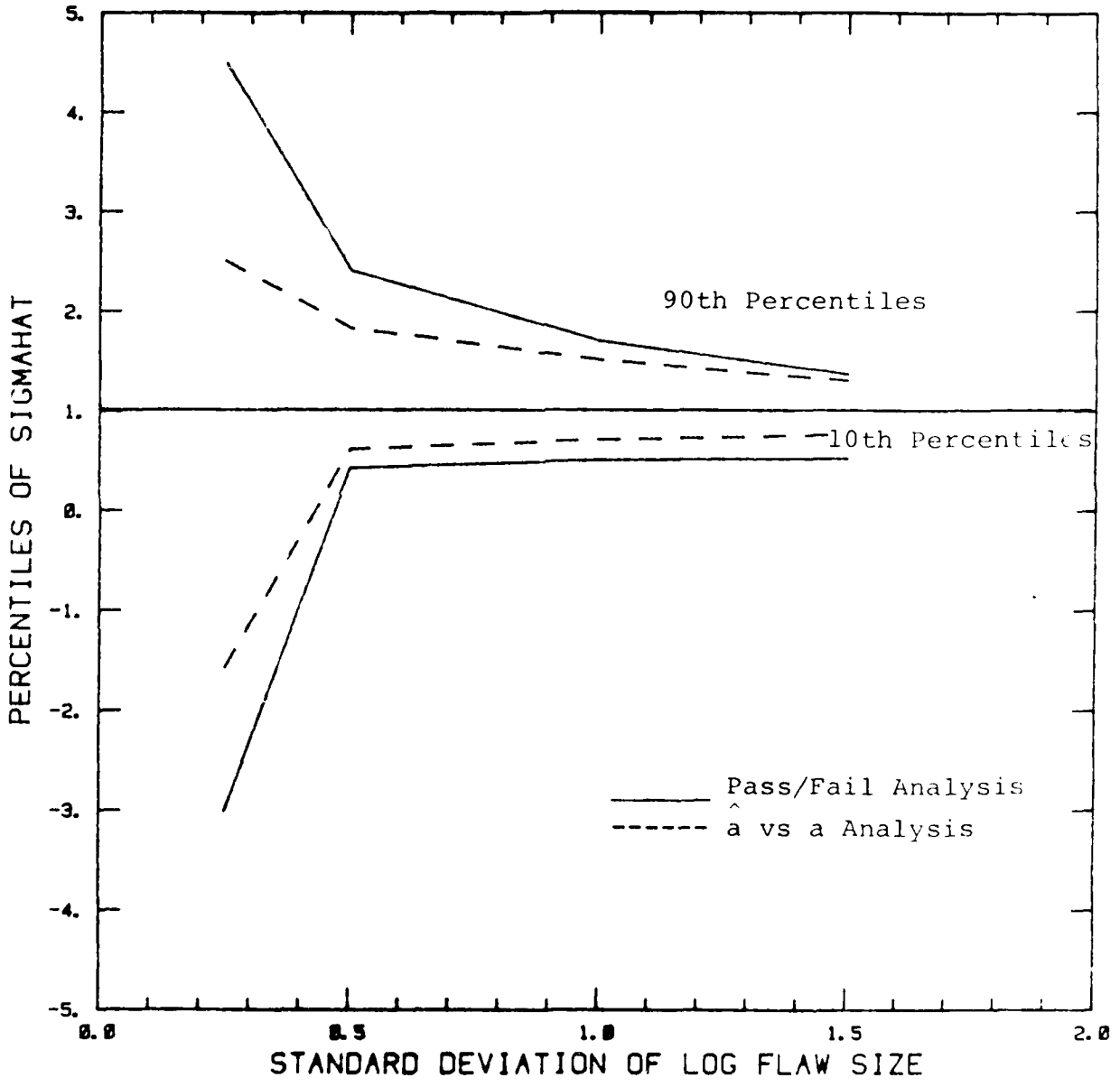
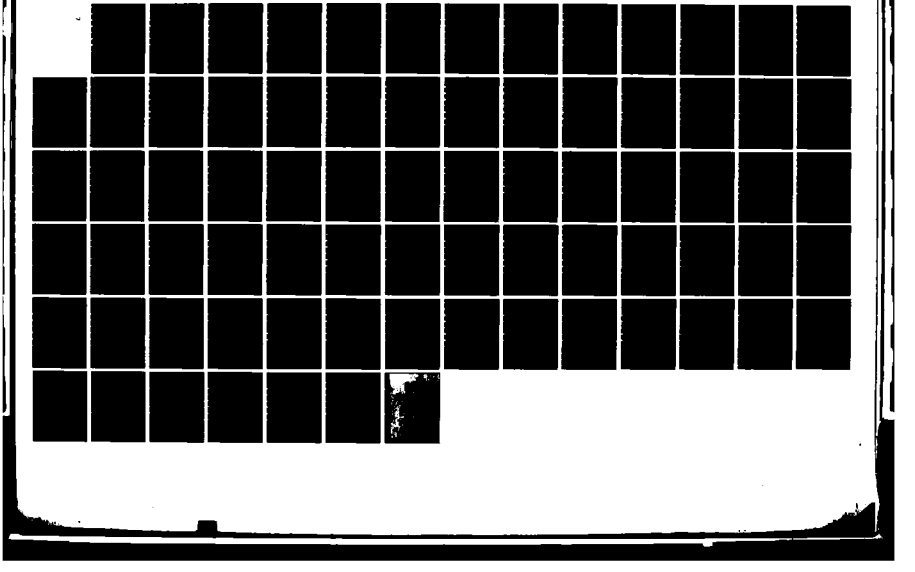
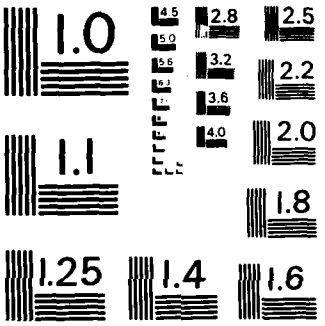


Figure 42. Tenth and 90th Percentiles of Estimates of σ as a Function of Standard Deviation of Log Flaw Sizes With Average Log Flaw Size of 0.5 and Sample Size of 40 - POD σ= 1.

AD-A142 001 FLAW DETECTION RELIABILITY CRITERIA VOLUME 1 METHODS AND RESULTS(U) DAYTON UNIV OH RESEARCH INST A P BERENS ET AL. APR 84 UDR-TR-83-137-VOL-1 2/2

UNCLASSIFIED AFWAL-TR-84-4022-VOL-1 F33615-82-C-5030 F/G 14/2 NL





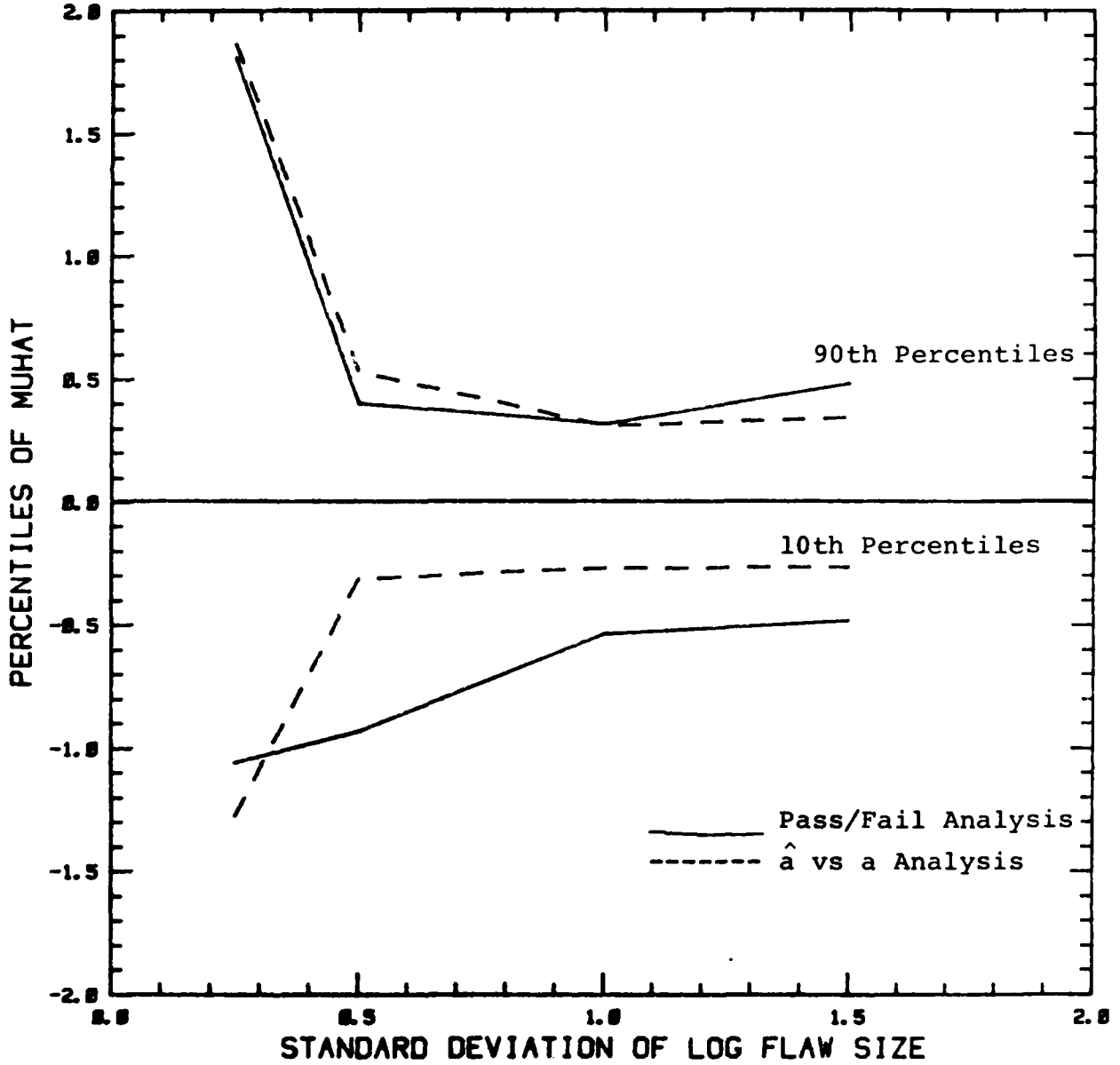


Figure 43. Tenth and 90th Percentiles of Estimates of μ as a Function of Standard Deviation of Log Flaw Sizes With Average Log Flaw Size of 2.0 and Sample Size of 30 - POD $\sigma = 1$.

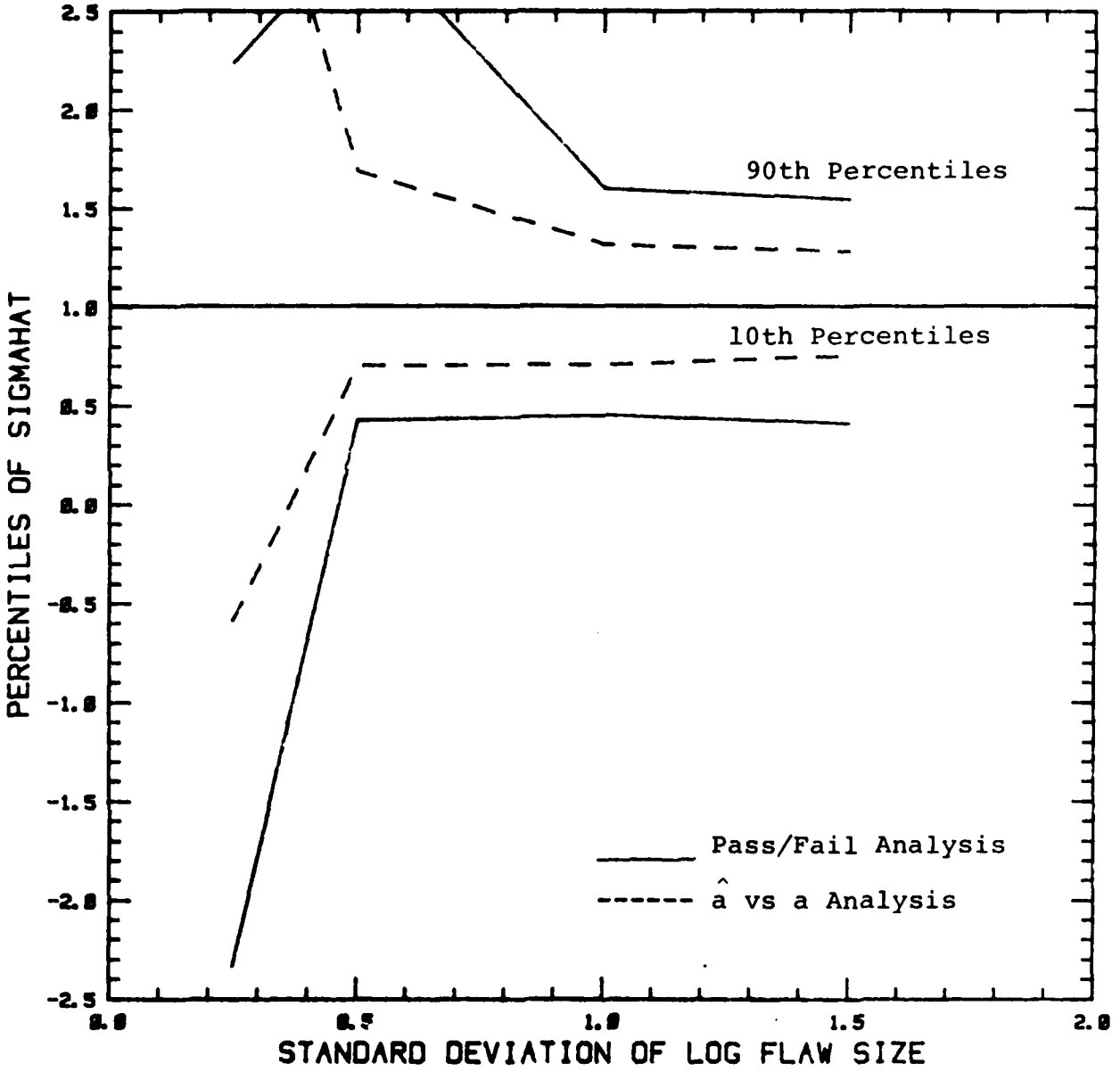


Figure 44. **Tenth and 90th Percentiles of Estimates of σ as a Function of Standard Deviation of Log Flaw Sizes With Average Log Flaw Size of 2.0 and Sample Size of 30 - POD σ = 1.**

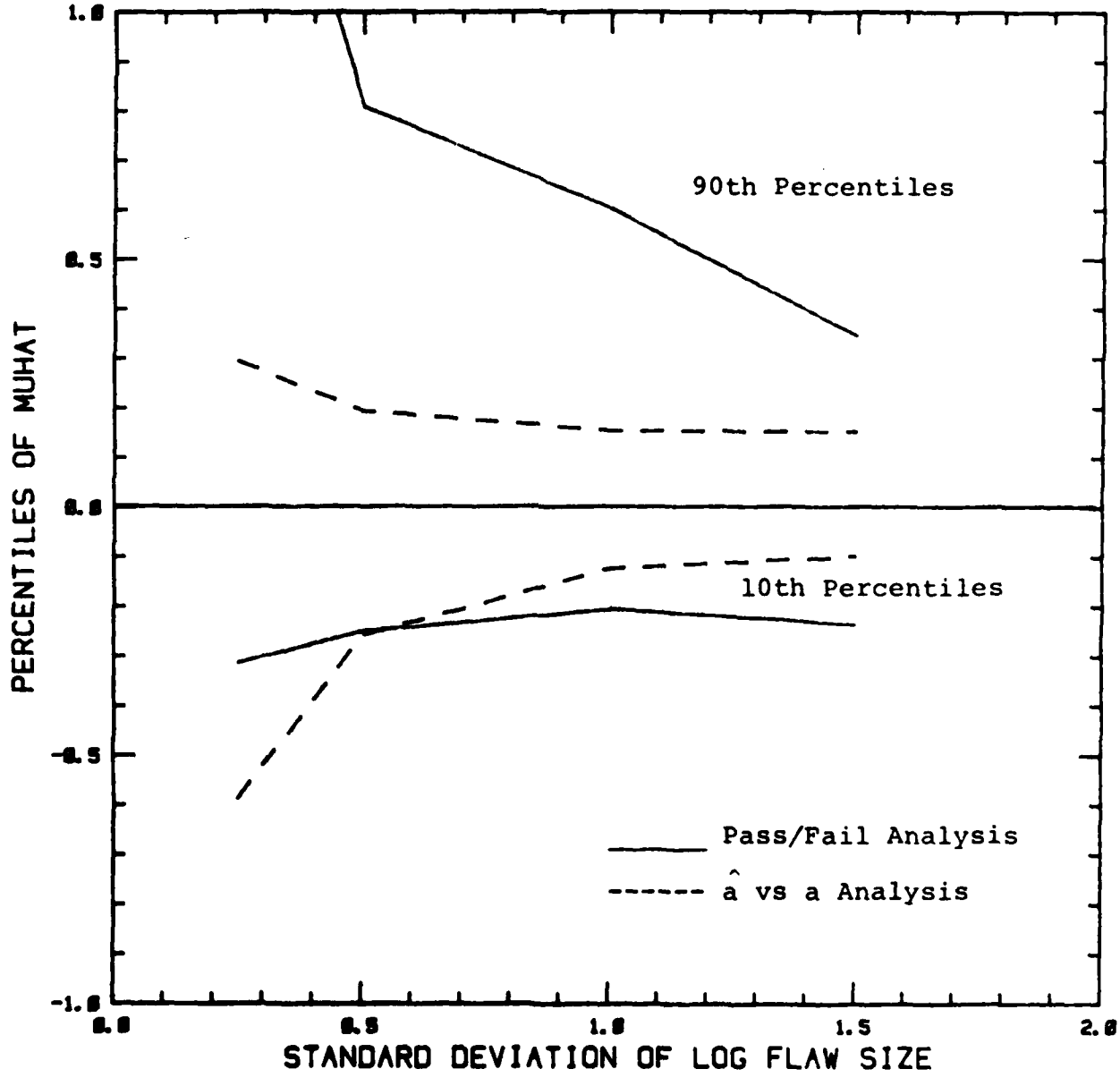


Figure 45. Tenth and 90th Percentiles of Estimates of μ as a Function of Standard Deviation of Log Flaw Sizes With Average Log Flaw Size of 0.5 and Sample Size of 100 - POD $\sigma = 1$.

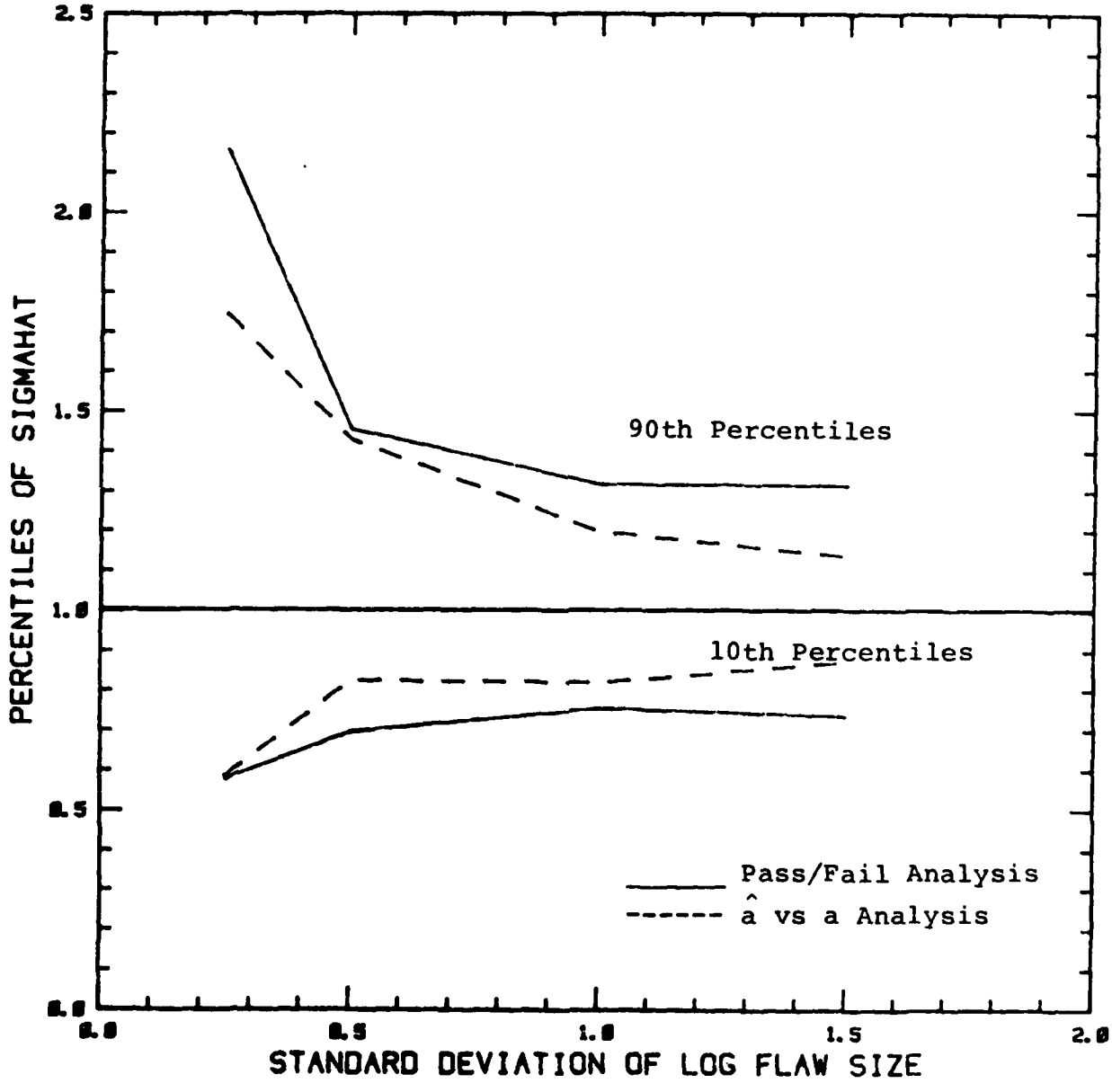


Figure 46. Tenth and 90th Percentiles of Estimates of σ as a Function of Standard Deviation of Log Flaw Sizes With Average Log Flaw Size of 0.5 and Sample Size of 100 - POD σ = 1.

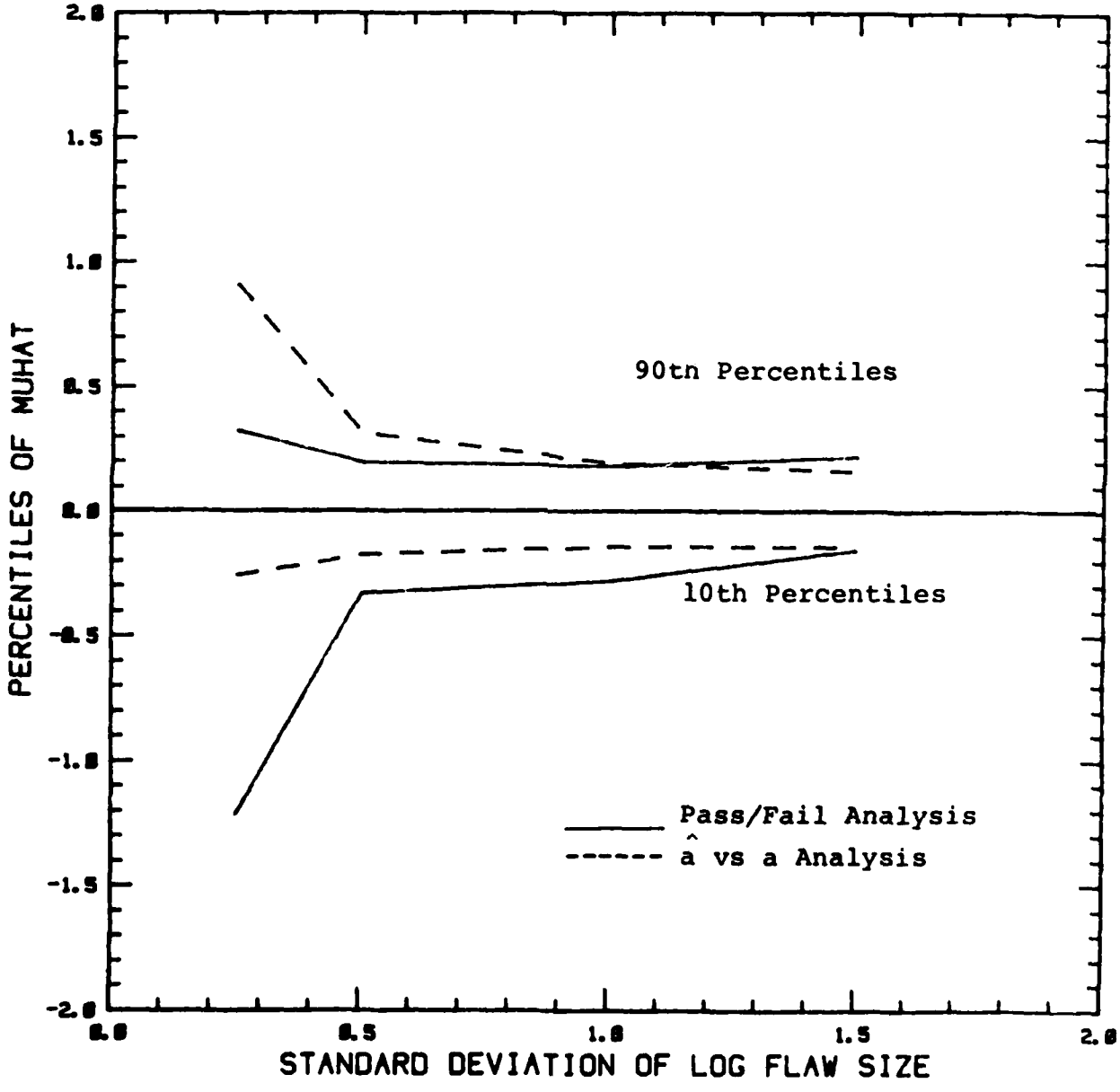


Figure 47. Tenth and 90th Percentiles of Estimates of μ as a Function of Standard Deviation of Log Flaw Sizes With Average Log Flaw Size of 2.0 and Sample Size of 100 - POD $\sigma = 1$.

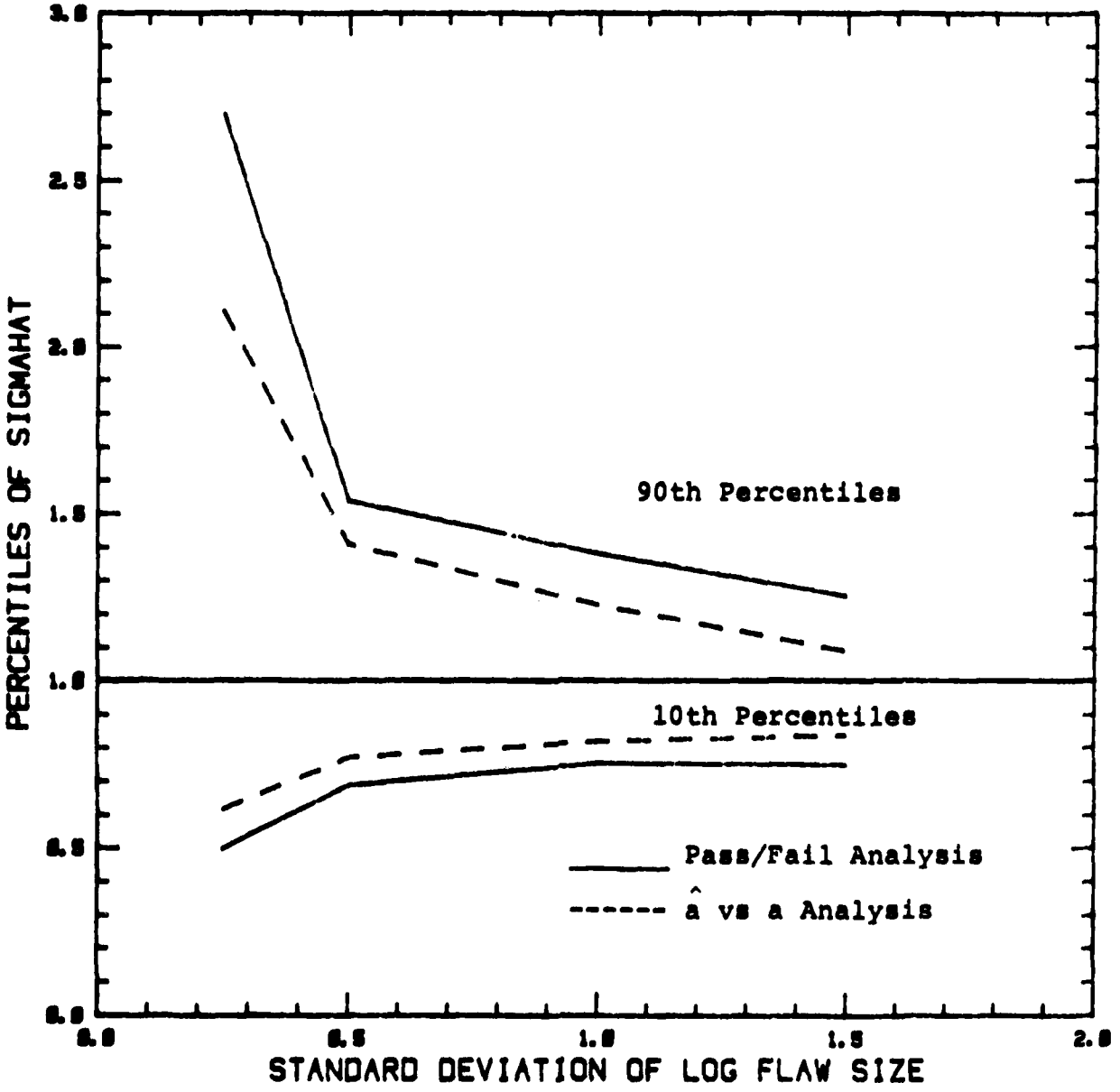


Figure 48. Tenth and 90th Percentiles of Estimates of σ as a Function of Standard Deviation of Log Flaw Sizes With Average Log Flaw Size of 2.0 and Sample Size of 100 - POD σ = 1.

The ability of an estimator to repeat the same value or close to the same value is called precision. The precision is usually measured by the variance or standard deviation of the estimator. Biased estimators pose a unique problem in defining precision. The variance and standard deviation measure the scatter in a random variable about its mean; which in a biased estimator is not the value of interest. The scatter about the parameter being estimated is more pertinent to the ability of the estimator to estimate the parameter.

The mean square error (MSE) is the measure of the scatter of an estimator about the parameter it is estimating. The MSE is given by

$$MSE = E(\ \hat{\theta} - \theta\ )^2 = Var\ \hat{\theta} + b^2 \qquad (36)$$

The mean square error can be used to compare two analysis methods regardless of any biasses.

The relative efficiency is a standard statistical measure for comparing two analysis methods and is given by

$$RE = \frac{MSE\ (\hat{\vartheta}_2)}{MSE\ (\hat{\vartheta}_1)} \qquad (37)$$

which is the relative efficiency of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ Tables 3, 4, and 5 give the MSE's for both analysis methods and the RE's of pass/fail analysis with respect to $\hat{a}$ versus a analysis for μ and σ. Relative efficiencies can be interpreted in terms of the sample sizes required of the two analysis methods to produce the same precision. The sample size required for $\hat{\theta}_2$ to produce the same precision as $\hat{\theta}_1$ is RE times the sample size for $\hat{\theta}_1$. If RE is less than 1, $\hat{\theta}_2$ requires a smaller sample than $\hat{\theta}_1$ and therefore is a better estimate.

TABLE 3. MEAN SQUARE ERRORS AND RELATIVE EFFICIENCIES FOR TWO ANALYSIS METHODS AND SELECTED VALUES OF δ, λ AND n - μ=0, σ=1.

δ	λ	n	$\hat{\mu}$ MSE_1^*	$\hat{\mu}$ MSE_2^{**}	RE	$\hat{\sigma}$ MSE_1^*	$\hat{\sigma}$ MSE_2^{**}	RE
ln(.5)	0.25	30	4268	7.11	0.017	5538	10.9	0.0020
		100	113.2	1.48	0.013	148	3.03	0.020
	0.5	30	1.93	1.40	0.73	9.79	3.02	0.31
		100	0.11	0.038	0.33	0.21	0.073	0.35
	1.0	30	0.16	0.069	0.44	0.28	0.08	0.30
		100	0.024	0.015	0.63	0.058	0.018	0.32
	1.5	30	0.11	0.049	0.44	0.18	0.054	0.30
		100	0.029	0.0088	0.31	0.046	0.012	0.25
ln(2)	0.25	30	32.05	96.80	3.02	116	256	2.19
		100	8.50	0.55	0.064	24.67	1.16	0.047
	0.5	30	5.06	0.11	0.022	7.41	0.28	0.038
		100	0.088	0.044	0.50	0.13	0.061	0.48
	1.0	30	0.11	0.049	0.43	0.27	0.068	0.25
		100	0.030	0.020	0.68	0.058	0.027	0.46
	1.5	30	0.13	0.057	0.43	0.22	0.046	0.20
		100	0.021	0.013	0.60	0.037	0.014	0.38

*Subscript 1 denotes pass/fail analysis

**Subscript 2 denotes $\hat{a}$ vs a analysis

TABLE 4. MEAN SQUARE ERRORS AND RELATIVE EFFICIENCY FOR TWO ANALYSIS METHODS AND SELECTED VALUES OF σ and n - $\mu=0$, $\delta=0$, $\lambda=0.5$

σ	n	MSE_1^*	MSE_2^{**}	RE	MSE_1^*	MSE_2^{**}	RE
0.1	10	0.015	0.0017	0.11	0.01	0.0009	0.090
	30	0.0026	4×10^{-4}	0.15	0.0045	0.0004	0.089
	100	7.3×10^{-4}	1×10^{-4}	0.14	7.2×10^{-4}	1×10^{-4}	0.14
	400	1.7×10^{-4}	1.6×10^{-5}	0.09	2×10^{-4}	3×10^{-5}	0.13
0.25	10	0.026	0.01	0.38	0.043	0.0082	0.19
	30	0.0081	0.0025	0.31	0.012	0.0025	0.21
	100	0.0019	0.00083	0.43	0.0026	6.3×10^{-4}	0.24
	400	4.8×10^{-4}	1.4×10^{-4}	0.30	7.3×10^{-4}	2×10^{-4}	0.27
0.5	10	0.27	0.12	0.45	15,376	12,769	0.83
	30	0.012	0.010	0.83	0.036	0.017	0.46
	100	0.0049	0.0024	0.49	0.012	0.0054	0.46
	400	0.0011	5.8×10^{-4}	0.53	0.0022	0.0010	0.45
1.0	10	139.8	0.22	0.0016	59,049	36,100	0.61
	30	1.21	0.078	0.065	189	2.50	0.0026
	100	0.020	0.012	0.62	0.12	0.12	1.02
	400	0.0035	0.0025	0.72	0.018	0.0096	0.53
1.5	10	7.84	53.72	6.85	4×10^6	844,563	0.21
	30	0.58	44.3	76.2	8.05	784	97.3
	100	0.044	0.033	0.74	4.51	0.54	0.12
	400	0.0076	0.062	0.82	0.097	0.058	0.60

* Subscript 1 denotes pass/fail analysis

** Subscript 2 denotes $\hat{a}$ vs a analysis

TABLE 5. MEAN SQUARE ERRORS AND RELATIVE EFFICIENCY FOR TWO ANALYSIS METHODS AND SELECTED VALUES OF δ, λ, σ, and $n-\mu=0$.

δ	λ	σ	n	$\hat{\mu}$			$\hat{\sigma}$		
				MSE_1^*	MSE_2^{**}	RE	MSE_1^*	MSE_2^{**}	RE
0	0.25	0.25	10	0.039	2.8	74	0.51	76.6	151
			30	0.0045	0.0028	0.62	0.013	0.004	0.31
			100	0.0013	0.0007	0.55	0.0028	0.0014	0.49
			400	0.0003	0.0002	0.57	0.00055	0.0003	0.58
0	1.0	1.0	10	0.43	16.52	38	5.66	1711	302
			30	0.086	0.047	0.54	0.28	0.14	0.48
			100	0.019	0.013	0.69	0.041	0.018	0.46
			400	0.0051	0.0028	0.55	0.010	0.0042	0.40
ln(0.5)	0.5	0.5	10	1228	1.22	0.001	8.57	1.21	0.14
			30	1.62	0.040	0.024	0.18	0.019	0.11
			100	0.017	0.0084	0.50	0.015	0.0062	0.41
			400	0.0035	0.0025	0.70	0.0039	0.0013	0.33

*Subscript 1 denotes pass/fail analysis

**Subscript 2 denotes $\hat{a}$ vs a analysis

All the simulations summarized in Table 4 used a log-normal crack length distribution with $\delta=0$ and $\lambda=0.5$ for the mean and standard deviation of the log crack length. Sample sizes of 10, 30, 100 and 400 were used with the POD parameter $\sigma=0.1$, 0.25, 0.5, 1.0 and 1.5. The mean square errors increase with σ, implying that steeper POD functions are easier to estimate. The relative efficiencies of $\hat{\mu}$ and $\hat{\sigma}$ appear to increase with σ, but are relatively unaffected by sample size.

The relative efficiencies are generally less than 1 implying that the $\hat{a}$ versus a analysis is generally better than the pass/fail analysis. The increasing relative efficiency the pass/fail analysis with σ is a result of the change in [illegible]read of the flaw sizes with respect to the spread of the POD fun[illegible]on. For small σ, the spread in flaw sizes is much larger than t[illegible] spread in the POD function, resulting in more flaws in the s[illegible]ple with POD's close to 0 and 1. In a pass/fail analysis, flaws with POD's close to 0 or 1 supply practically no information about the shape of the POD curve. The effective sample size is decreased because the flaws in the tails do not supply as much information as the flaws in the middle.

In the $\hat{a}$ versus a analysis, the shape of the POD function is estimated from the distribution of deviations from the mean. All flaws in the sample supply the same information about the distribution of deviations; therefore, the effective sample size is not diminished by flaws in the tails of the POD function. The result is that the $\hat{a}$ versus a analysis is much more efficient than the pass/fail analysis when the spread in flaw sizes is larger than the spread in the POD function.

The results of simulations in which the spread in flaw sizes were equal to the spread in the POD function are summarized in Table 4. In two cases, the median of the flaw sizes was 1 and in the third, the median flaw size was 0.5 to investigate the effect of shifting flaw sizes away from the median crack detection length. For $\sigma = 0.25$ and 1.0 the relative efficiencies are about the same as for $\sigma = \lambda = 0.5$ in Table 4.

SECTION 4
EVALUATION OF POD CHARACTERISTICS

The link between NDI reliability and risk of structural failure is the exceedance probability of having a flaw greater than the initial size (a_o or a_{NDE}) assumed to be present at the beginning of a usage period. Given a particular inspection system and a population of details, this exceedance probability $H(a_{NDE})$ is constant regardless of the method for selecting a_{NDE}. However, there is uncertainty in the estimation of the POD function and the characteristics of the crack size distributions are not yet predictable.

The analyses of this section were made to evaluate the effects of crack size distribution assumptions and variation in POD parameter estimates on measures of inspection efficacy. First, it is assumed that flaw sizes are modeled by the growth of Weibully distributed equivalent initial cracks. Then, the rogue flaw models will be analyzed. These results will be contrasted with those of the equivalent flaw distribution approach. All of the calculations of this section were made in terms of normalized crack lengths. Therefore, for notational simplicity, normalized crack lengths are designated as a (rather than a').

4.1 POD AND EQUIVALENT CRACK SIZE DISTRIBUTIONS

To evaluate the combined effect of inspection reliability and equivalent crack size assumptions, it was assumed that each structural detail contained an equivalent initial crack and that these cracks would grow due to fatigue mechanisms. In this study, it was assumed that the equivalent initial cracks were distributed in accordance with the Weibull distribution with a constant shape parameter of 1.5 and that the effect of the fatigue environment was measured by an increase in the median crack size. Figure 5 illustrates the five equivalent crack size distributions that were considered with median normalized crack lengths of $\tilde{a}$=0.1, 0.25, 0.5, 0.75 and 1.0. As noted previously, the equivalent crack distribution with a normalized median crack length of 0.1 (i.e. $\tilde{a}$=0.1) is somewhat representative of the sizes of equivalent initial cracks that have been calculated in past studies. The other four would represent

the distribution of cracks that could be present at various ages in the life of the structure.

For each of the 5 equivalent crack size distributions, the probability of having a crack longer than a and missing it at an inspection, H(a), was calculated for selected combinations of the parameters μ and σ of the log logistics model of the POD(a) function. First, the median detectable flaw length, $a_{0.5}$, was considered to be one ($\mu = \ell n\ a_{0.5} = 0$) in the normalized crack length scale and the sensitivity of H(a) to variations in median crack length, $\tilde{a}$, and σ was determined. Then, for selected values of σ and $\tilde{a}$, the sensitivity of H(a) to variations in μ was calculated.

4.1.1 Effect of Median Crack Length

Figures 49 and 50 present the probability of having a crack greater than a in a structural detail after an inspection for the 5 equivalent crack size distributions for σ = 0.5 and 1.0, respectively. This probability of exceedance, H(a), can also be interpreted as the proportion of inspection sites which will have cracks greater than a after all sites have been inspected.

Note first that a particular choice of an a_{NDE} value can be evaluated using the H(a) function. For example, if the median crack length is half the median detection capability ($\tilde{a}$=0.5) and the a_{NDE} value is twice the median detection capability, (a_{NDE}=2) than the proportion of cracks which will exceed a_{NDE} after the inspection is 0.0002 if the POD function has a σ=0.5 (Figure 49) and is 0.0008 (Figure 50) if the POD function has a σ=1.0. Conversely, the exceedance probability can be defined and the a_{NDE} to yield this exceedance probability can be determined. For example, again letting $\tilde{a}$=0.5, if H(a) is desired to be 0.0001, then a_{NDE} would be 2.15 times the median inspection capability if σ=0.5 or 2.4 times the median inspection capability if σ=1.0.

The effect of the median crack size on the exceedance probabilities is quite significant. Under this model for equivalent crack sizes and constant inspection capability, it is apparent that a single value of a_{NDE} throughout the life of the structure results

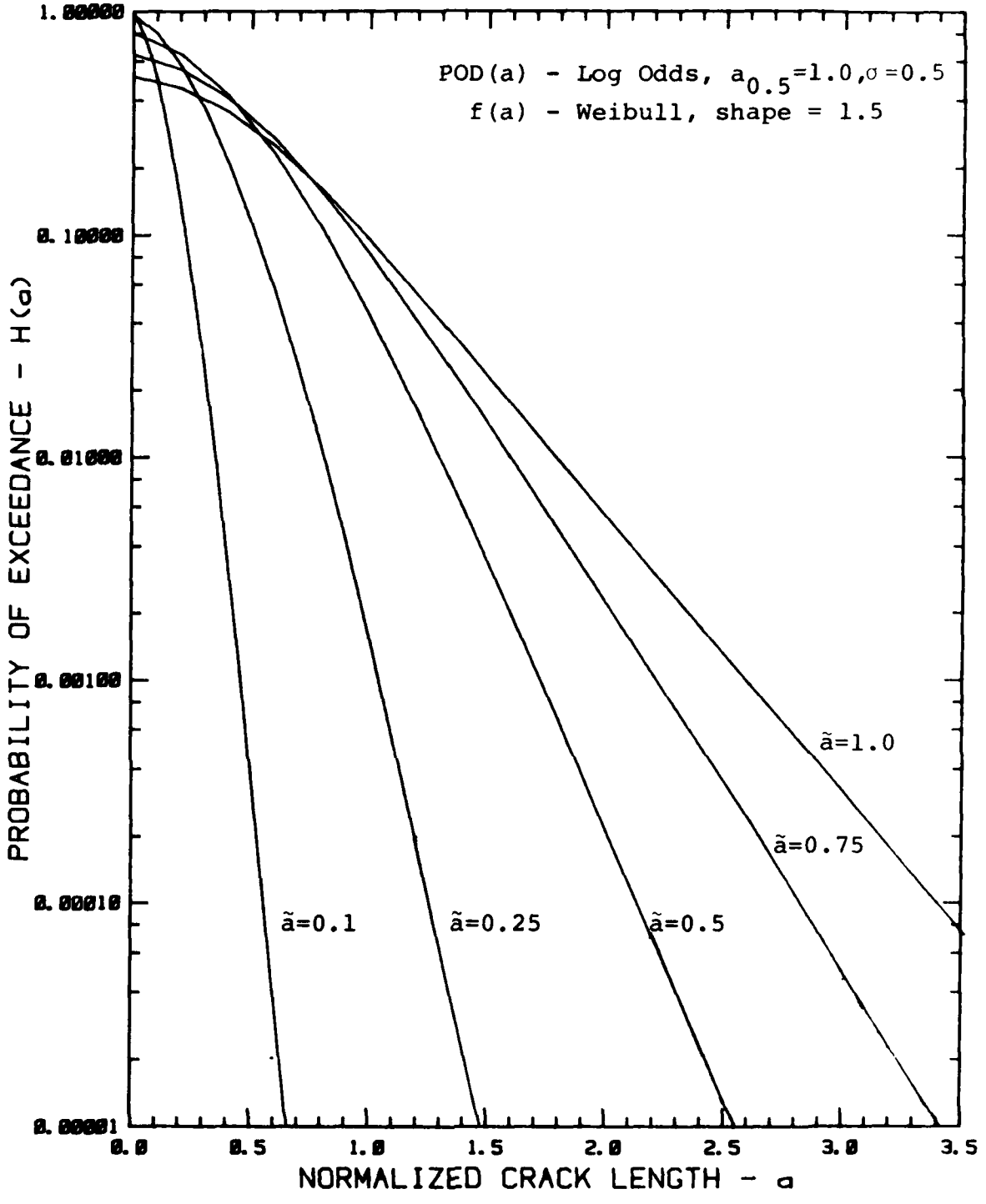


Figure 49. Exceedance Probabilities for Selected Median Crack Sizes - σ=0.5.

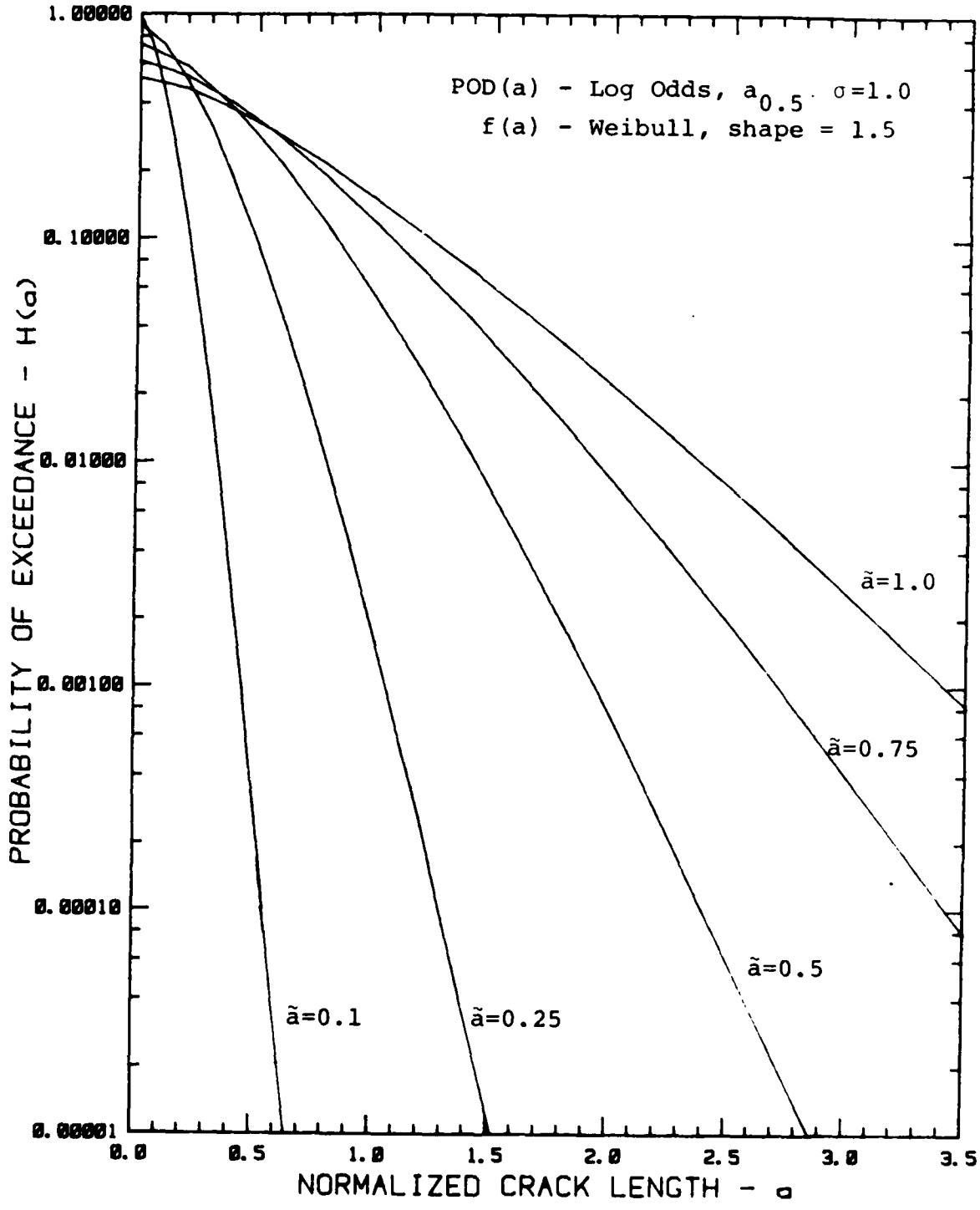


Figure 50. Exceedance Probabilities for Selected Median Crack Size - σ=1.0.

in orders of magnitude changes in the probabilities of having and missing flaws greater than a_{NDE} in the structure. If too large a value is selected, unneeded inspections would result. If too small a value is selected, the chances of structural failure increase significantly. As will be shown, for aircraft structure, it is not practical to have the median crack length of this model as large as half the median detection capability.

Most of the dependence of the exceedance probabilities on median crack length under this Weibull assumption is due to the fact that the H(a) calculation is being dominated by the crack size distribution. That is, the cracks are not long enough to be in the effective range of the inspection system. This can be seen in Figure 3 where the POD(a) function with σ=1.0 is less than 0.9 over the dominant range of crack lengths.

The effective inspection reliability,ER(a), for these crack length distributions are shown in Figures 51 and 52 for σ=0.5 and 1.0, respectively. The effective reliability is the proportion of all cracks greater than a which will be detected (equation (7)). As can be noted, these figures support the above conclusion that this inspection capability is relatively ineffective for the smaller crack size distributions. The curves for the three shorter crack length distributions were truncated at the crack length which is exceeded once in 10,000 details.

Figure 53 presents the percent of inspection sites which will be rejected as a function of median crack length for four σ values. Approximately half of all sites will be rejected when median crack length equals median detection capability regardless of σ. A relatively large proportion are rejected even when the median crack leng s half the median detection capability. Whether or not t ection proportions are too large would depend on the eco f the application.

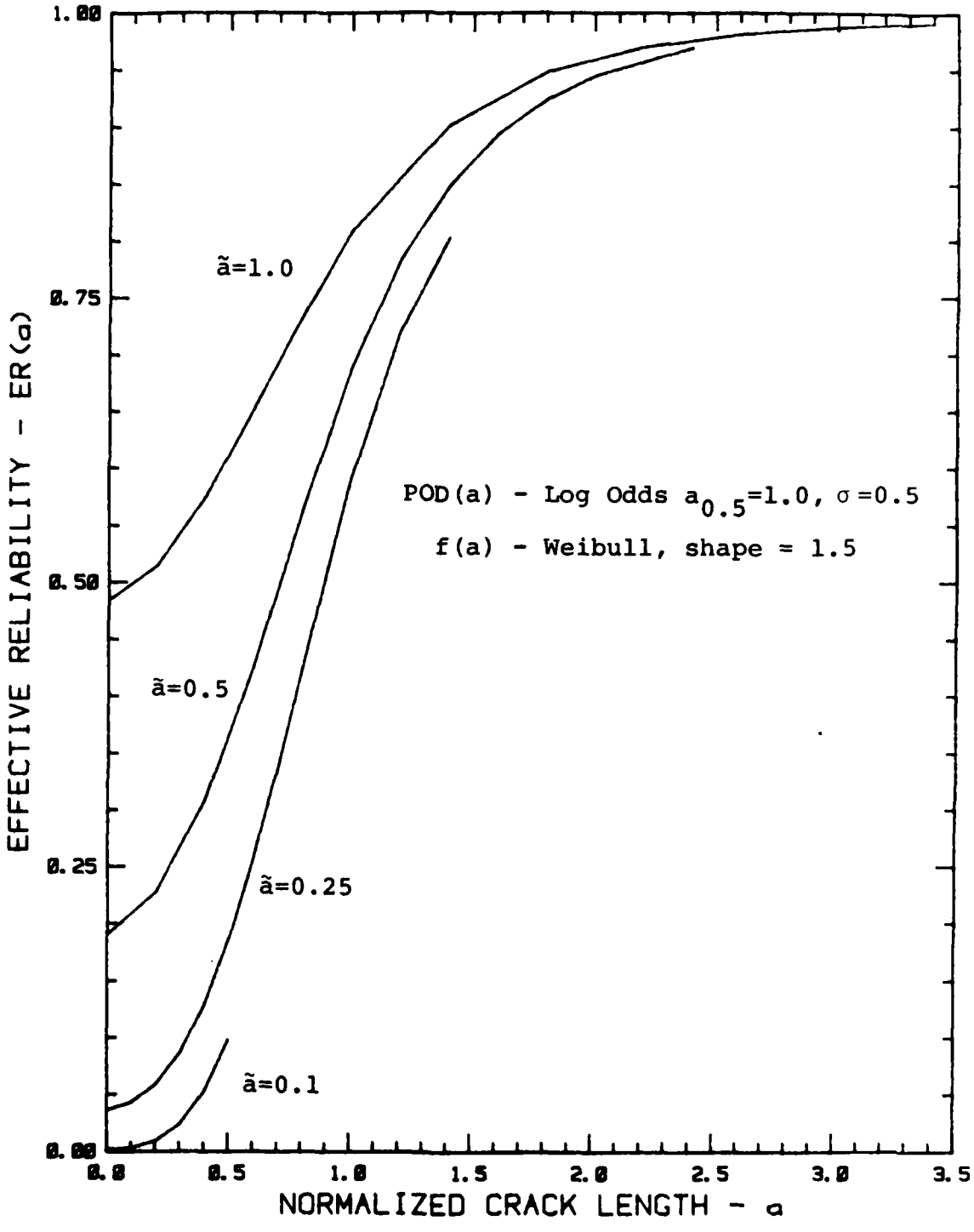


Figure 51. Effective Inspection Reliability for Selected Median Crack Sizes - σ=0.5.

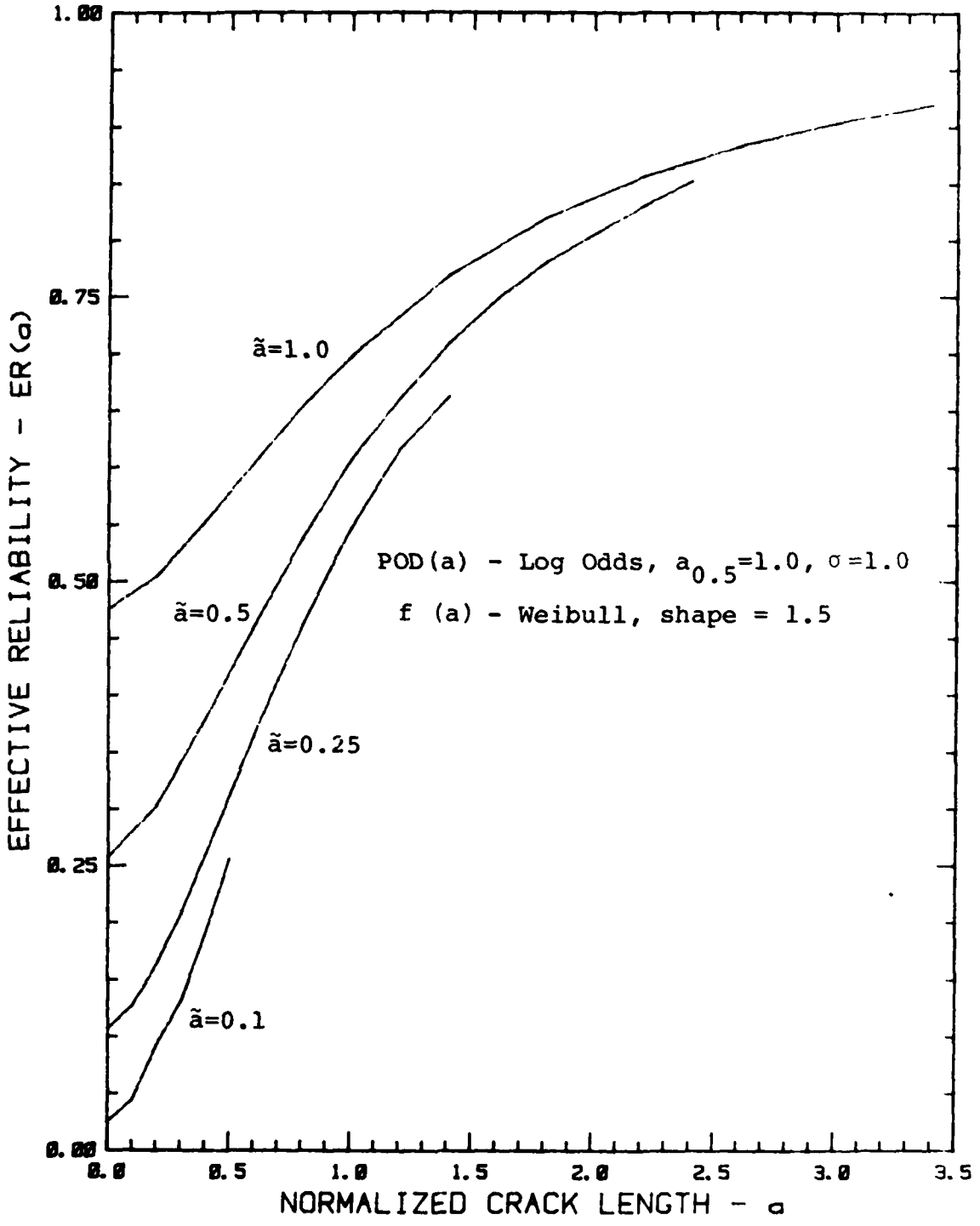


Figure 52. Effective Inspection Reliability for Selected Median Crack Sizes - σ=1.0.

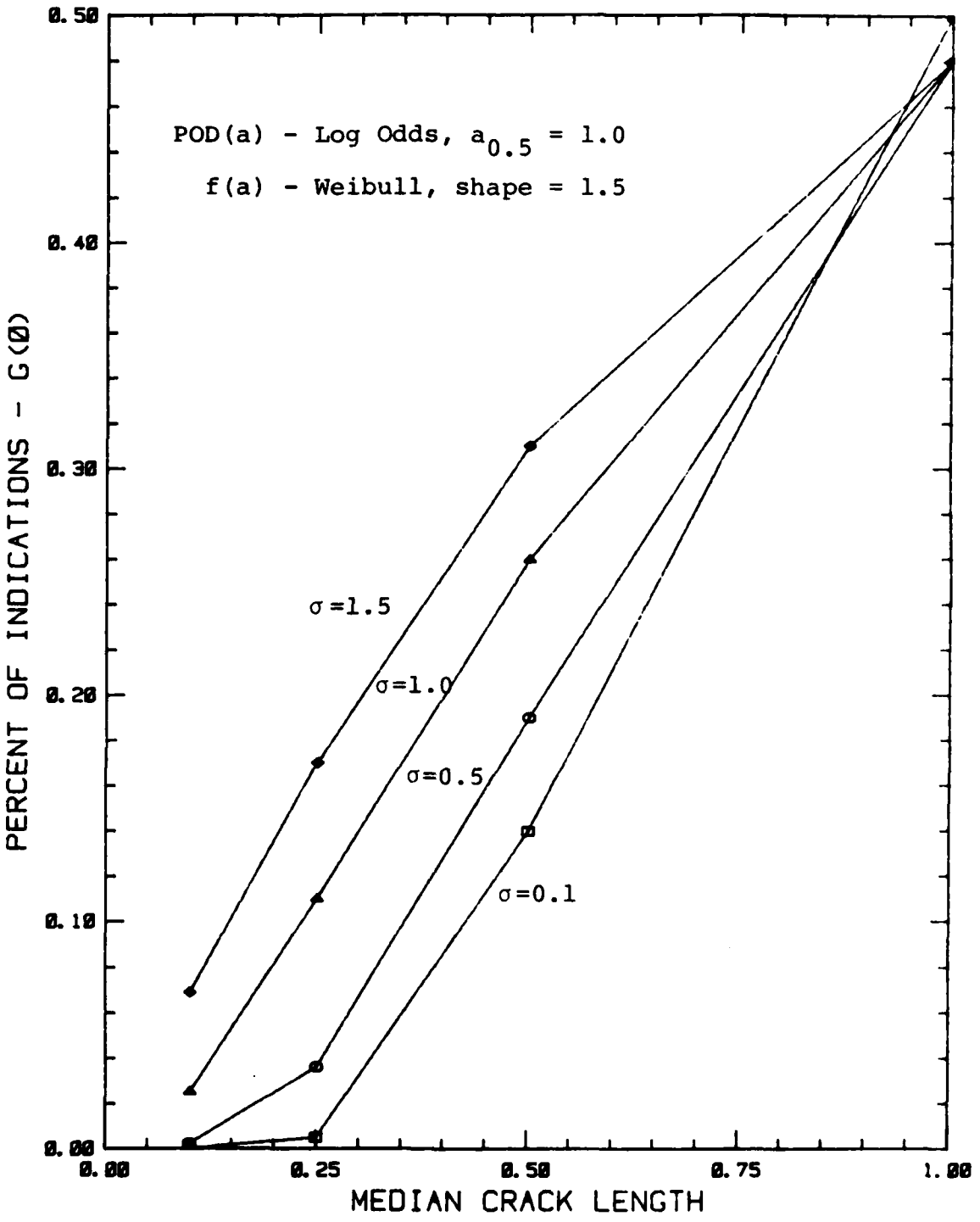


Figure 53. Proportion of Rejected Inspection Sites as a Function of Median Crack Size.

4.1.2 Effect of σ with Equivalent Crack Size Model

Figure 54 presents H(a) for σ=0.5 and 1.5 for the median crack lengths of 0.1, 0.25, and 0.5. This figure illustrates that variation in the flatness of the POD model (see Figure 4) has a relatively small effect on the exceedance probabilities. This can also be seen in the plots of Figure 55 which present inspection limits, as defined by the crack lengths at which H(a) reaches 0.00001, as a function of median crack length. The detection limits so defined are insensitive to POD shape for the smaller crack size distributions. Again, this result is due in large part to the lack of effectiveness of the inspection for the bulk of the inspection sites. Figure 56 presents the effective reliability for the four POD shapes when the median crack size is one fourth the median detection capability ($\tilde{a}$ = 0.25)

4.1.3 Effect of Median Detection Capability with Equivalent Crack Size Model

Figure 57 displays the log odds POD model for five values of median detectability ($a_{0.5}$ = 0.5, 0.75, 1.0, 1.25 and 1.5) with σ=0.5. H(a) was calculated for each of these POD functions and for equivalent crack length medians of $\tilde{a}$=0.1, 0.25 and 0.50. These results are summarized in Figure 58 in which a pair of H(a) functions are presented for each median crack length. Each pair of H(a) function spans the range of median detection capabilities. The exceedance probabilities are rather insensitive to changes in median inspection capability. Figure 59 presents inspection limits defined in terms of the crack length corresponding to an exceedance probability 0.00001 as a function of median detection capability for each of 4 values of σ and a median crack length of 0.25. Except for the steepest POD function (σ=0.1), the inspection limits are relatively insensitive to median detection capability. However, the inspection limits are separating as the inspection capability approaches the median crack size. Figure 60 presents the effective inspection reliability for the 5 median detection capabilities with σ = 1.0 and $\tilde{a}$ = 0.25. Obviously, the effective reliability increases as the median detection capability decreases.

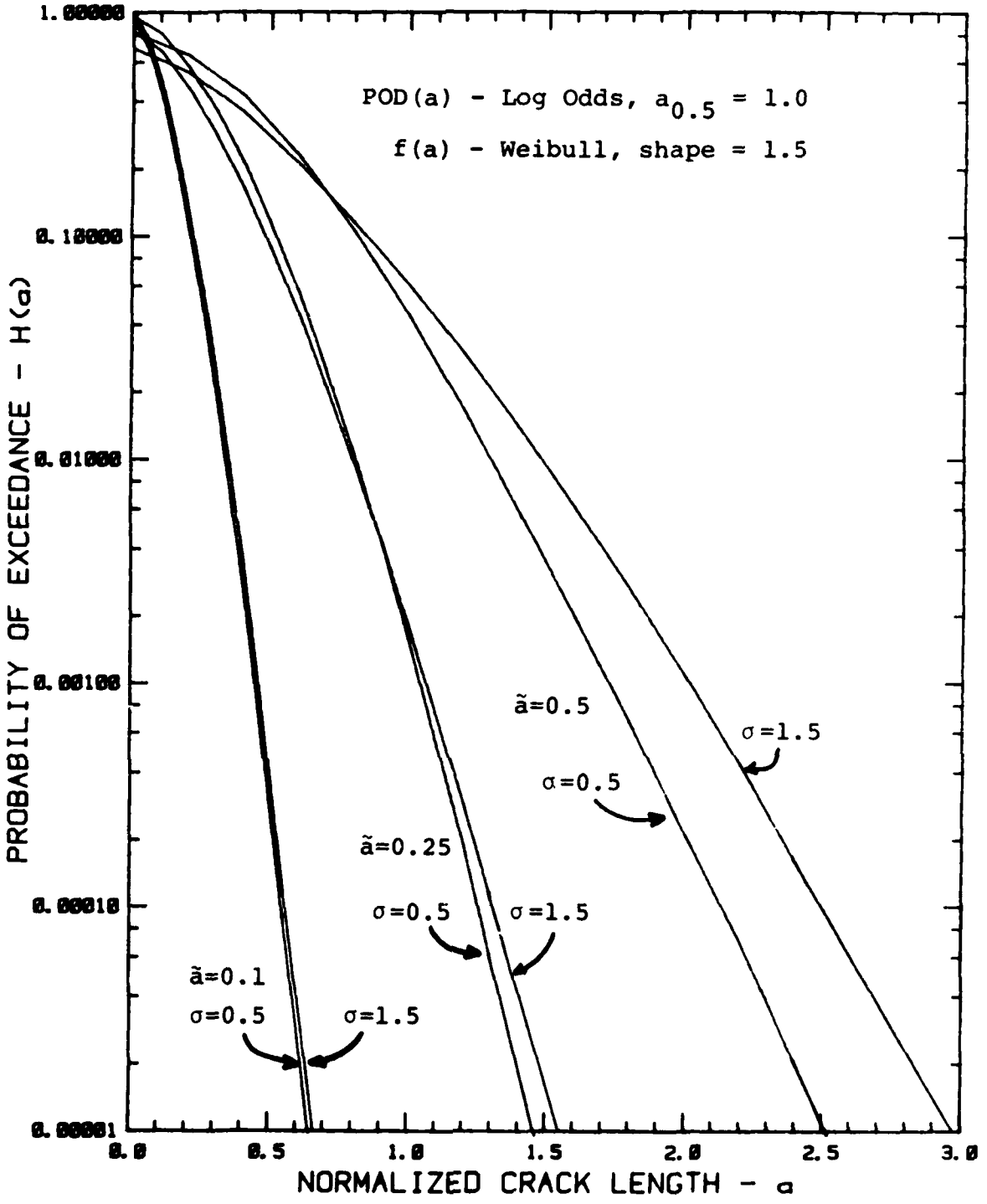


Figure 54. Exceedance Probabilities for Selected Median Crack Sizes and Ranges of σ.

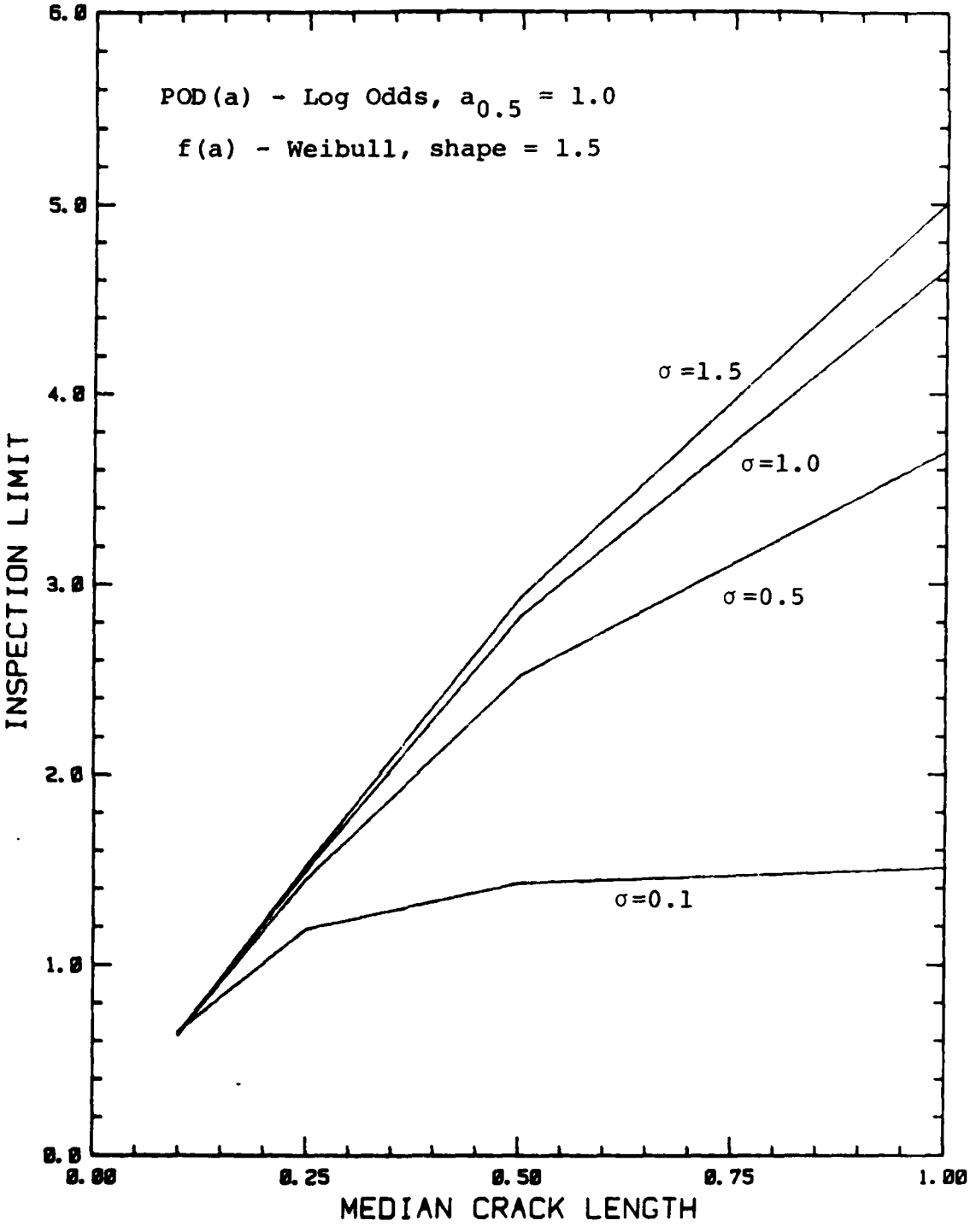


Figure 55. Inspection Limit Crack Length for H(a) = 0.00001 as a Function of Median Crack Length for Selected σ.

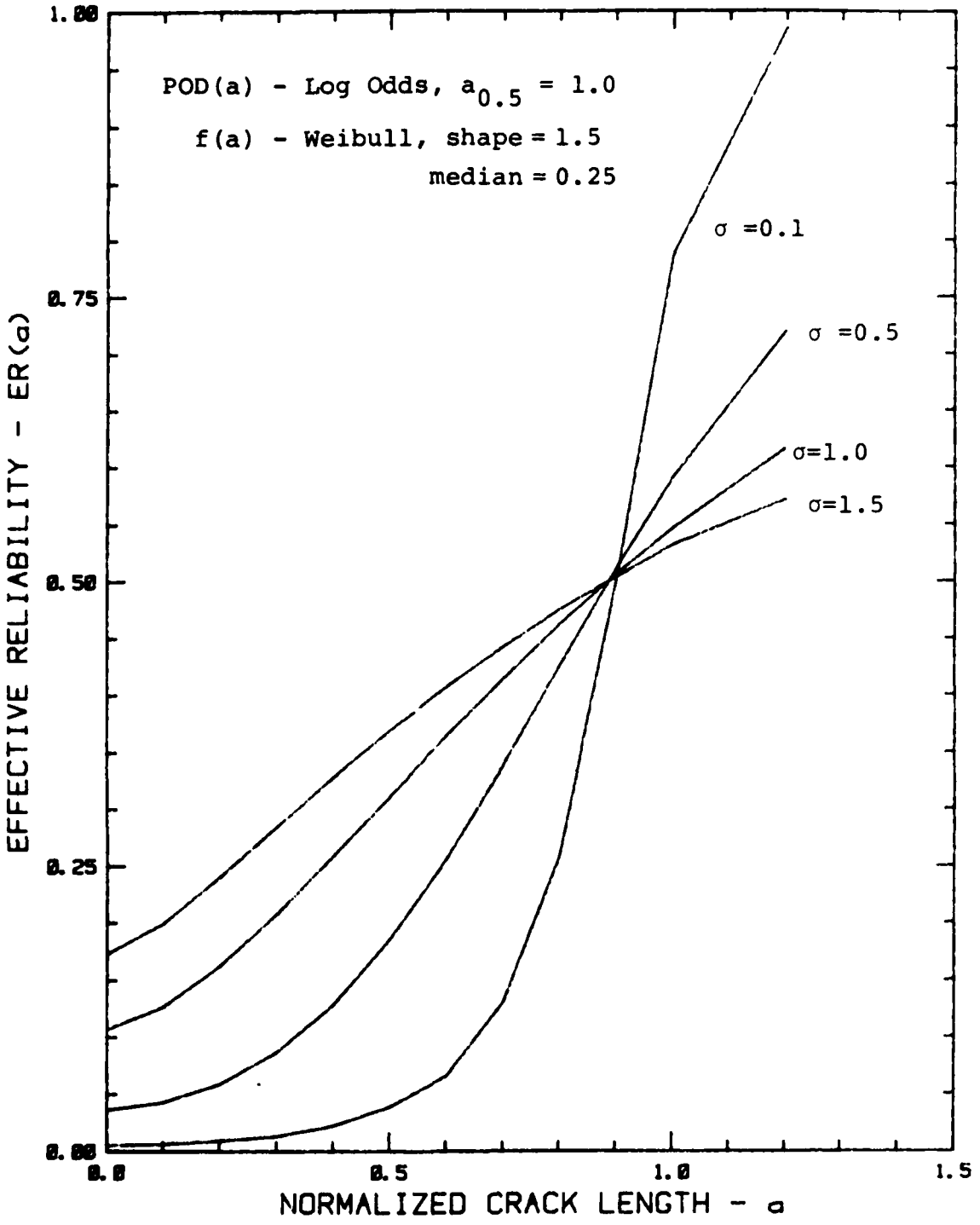


Figure 56. Effective Inspection Reliability for Selected σ Values - Median Crack Length = 0.25.

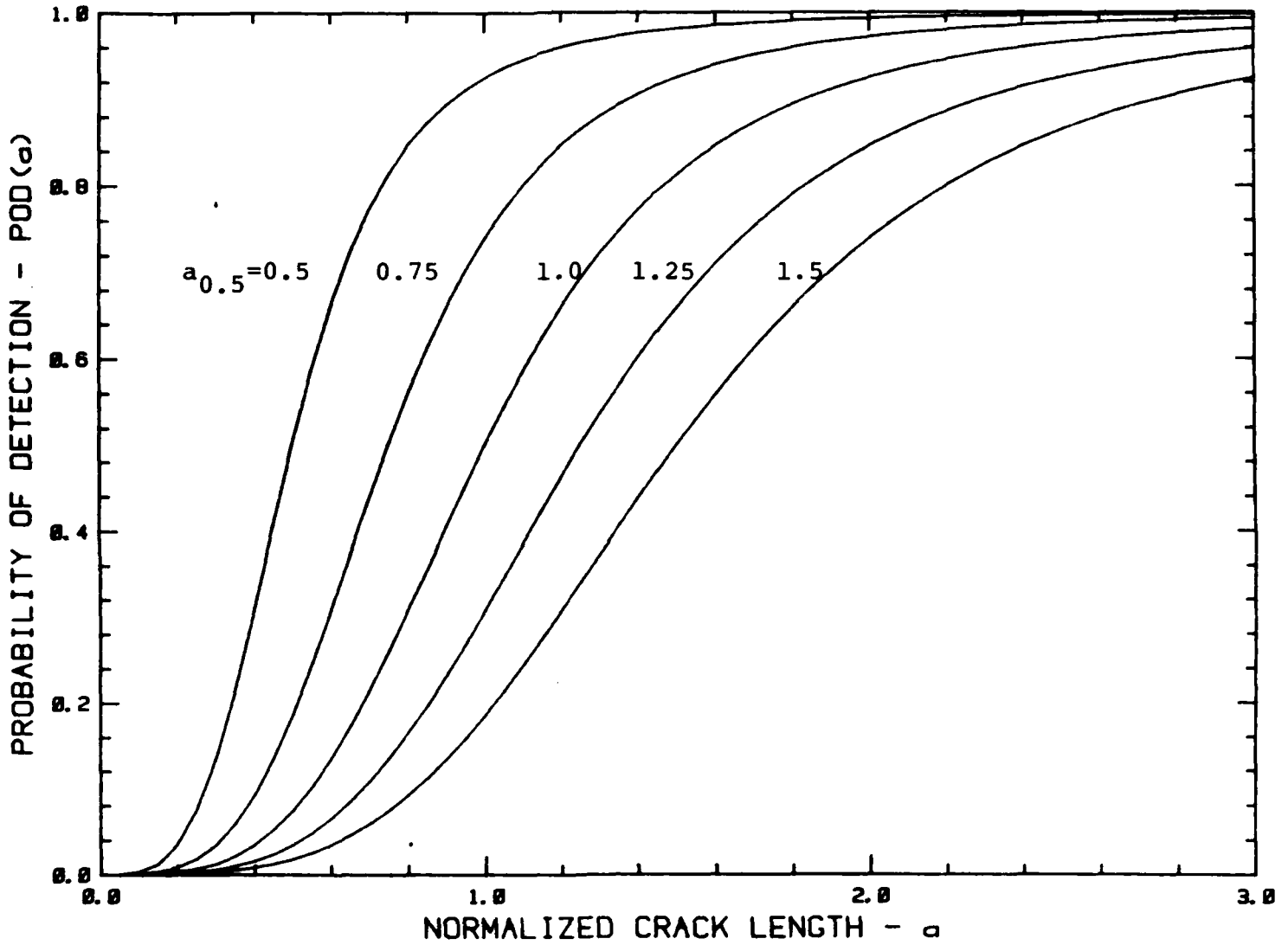


Figure 57. POD Functions for Selected Median Inspection Capabilities - σ=0.5

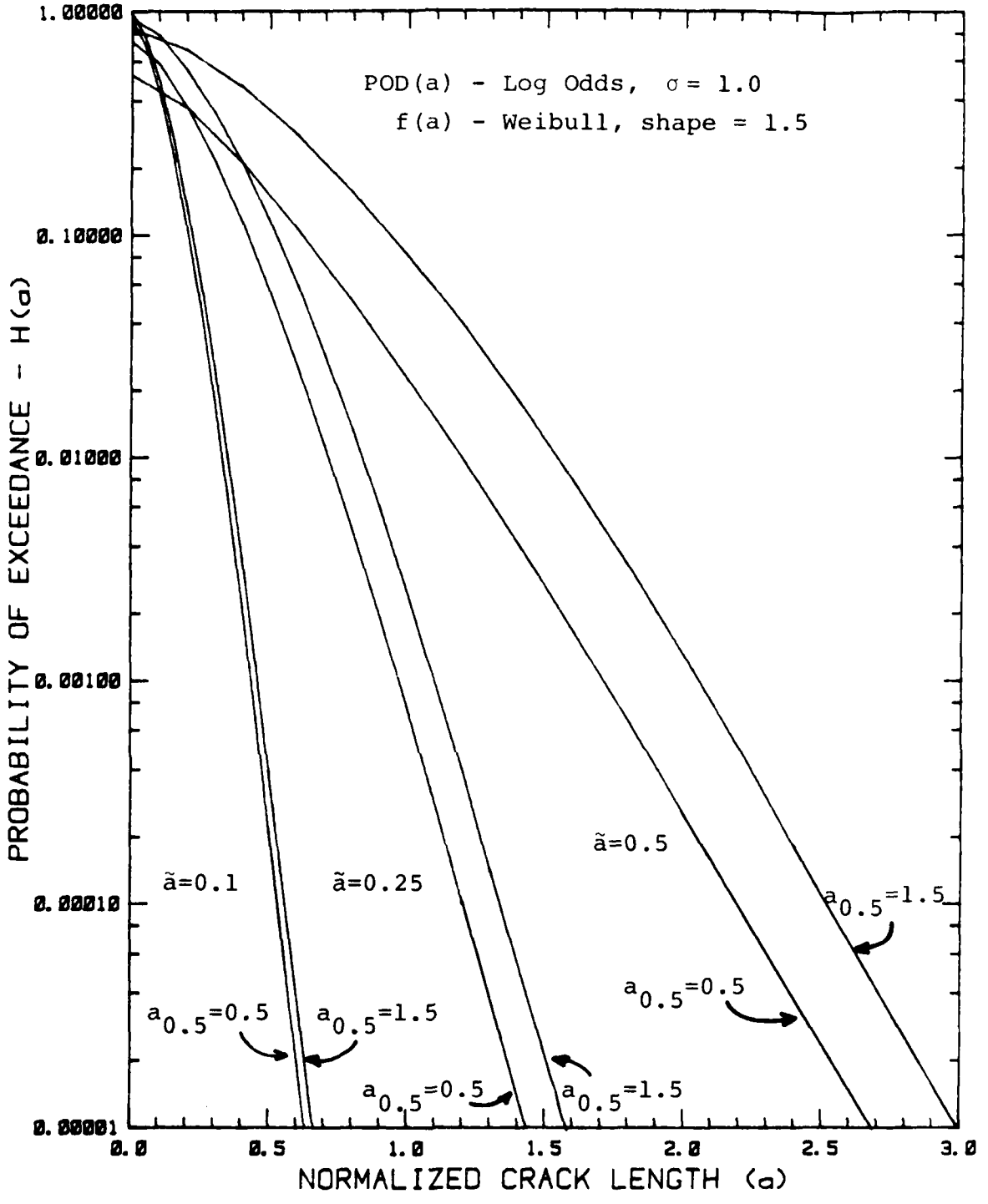


Figure 58. Exceedance Probabilities for Selected Median Crack Sizes and Ranges of Median Inspection Capabilities - σ=1.0.

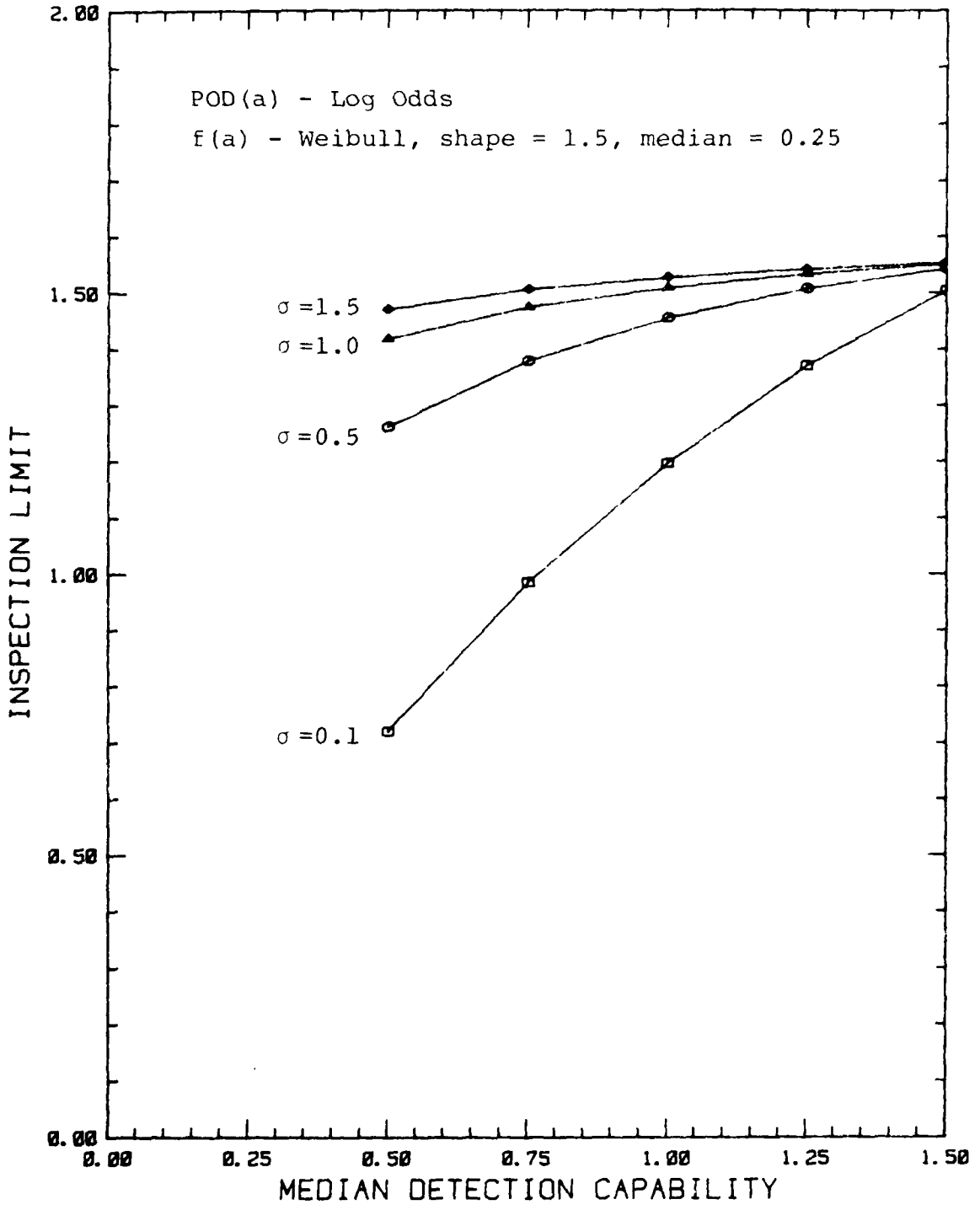


Figure 59. Inspection Limits for H(a) = 0.00001 as a Function of Median Detection Capability for Selected σ - Median Crack Length = 0.25.

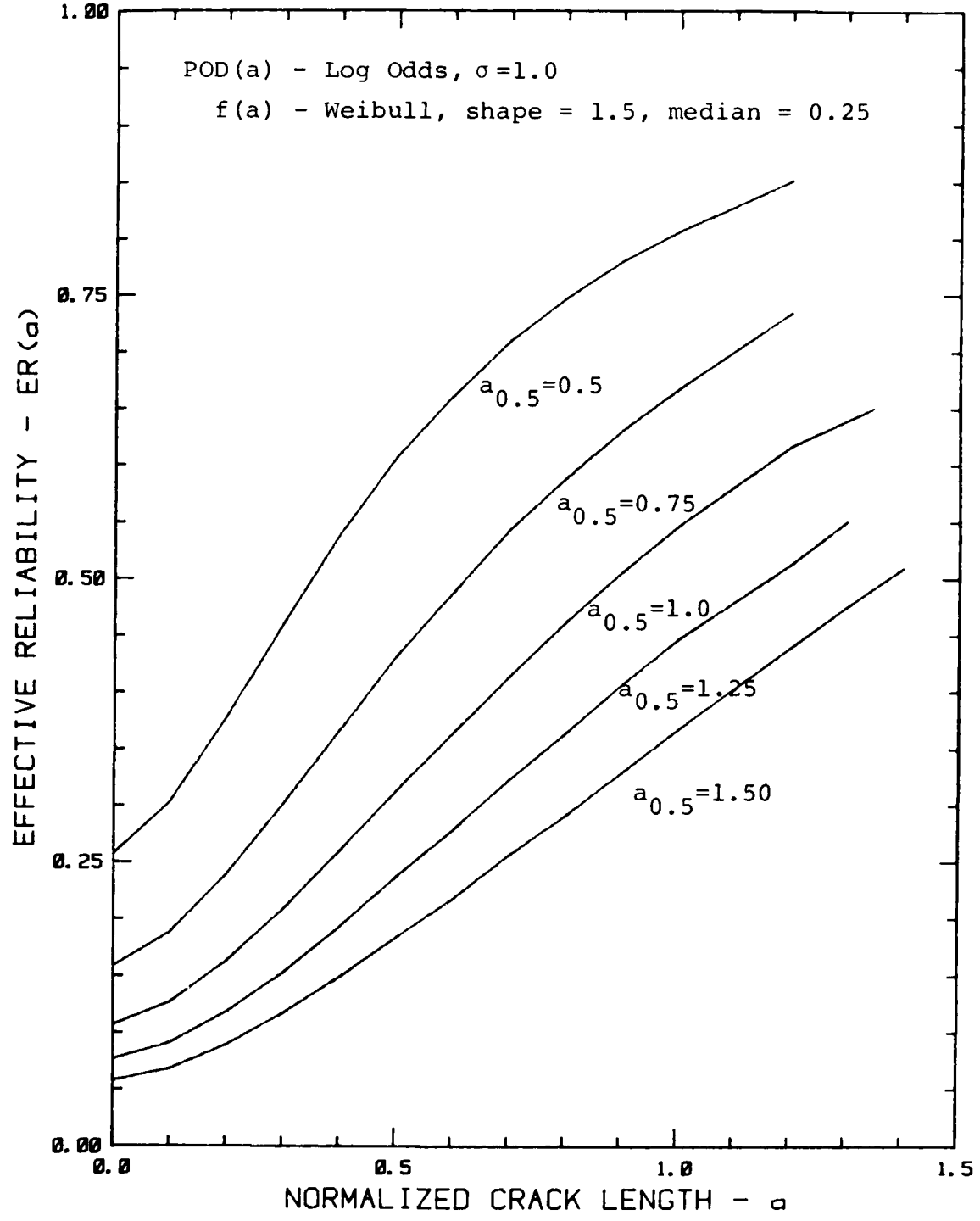


Figure 60. Effective Inspection Reliability for Selected Median Detection Capabilities - Median Crack Length = 0.25, σ=1.0.

4.2 POD AND ROGUE FLAW DISTRIBUTIONS

In the rogue flaw model for describing the flaw sizes in the population of structural details, it is assumed that only a very small proportion,p, of the details will contain a flaw. However, the sizes of these flaws can range up to some maximum value, A, for example, the critical crack size for the structure (as determined by the maximum expected stress level in the next usage period). When considering the exceedance probabilities under this rogue flaw model, p H(a) is the total proportion of details which will have a flaw greater than a after the inspection. In this subsection, the factor p will not be considered as it is assumed to be a constant (on the order of 0.001 or less).

In the following paragraphs, three rogue flaw distributions are compared in terms of their effect on exceedance probabilities and effective inspection reliabilities. It is concluded that for the purposes of this study, the uniform distribution provides an acceptable model. The uniform distribution is then used to evaluate the sensitivity of POD variations on exceedance probabilities and effective inspection reliabilities.

4.2.1 Comparison of Rogue Flaw Models

As discussed in Section 2, three families of distributions were considered as candidate rogue flaw models: a) the uniform distribution for which any flaw size between 0 and A is equally likely; b) the triangular distribution which makes smaller flaws more likely; and c) the exponential distribution truncated at A which, for small median crack lengths, makes small sizes more likely than the triangular distribution. The probability density functions for these distributions is shown in Figure 6.

The exceedance probabilities, H(a), for the three rogue flaw models are presented in Figure 61 for the POD models with a median detection capability of one and σ=0.5 and 1.0. For these calculations it was assumed that the maximum possible crack size,A, was 10 times the median detection capability. This assumption is evaluated for the uniform distribution in a later

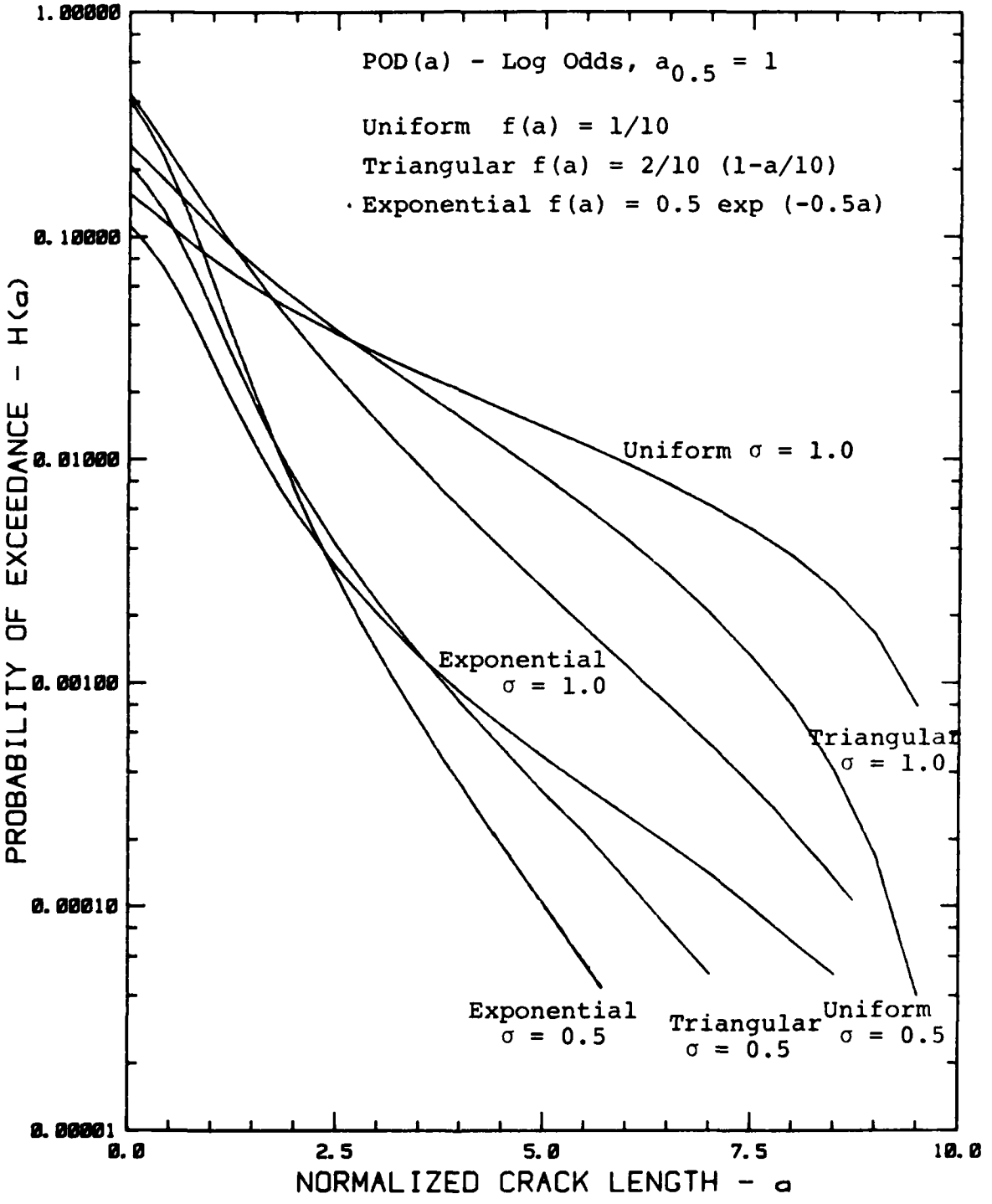


Figure 61. Exceedance Probabilities for Selected Rogue Flaw Crack Size Models - σ = 0.5 and 1.0.

paragraph. A median crack of 2 was selected for the exponential distribution as a compromise between over emphasis on small cracks and getting too close to the triangular density function. Note again that these H(a) values must be multiplied by the proportion of rogue flaws,p, in the population to obtain the total probability of exceedance.

The exceedance probabilities for the three models are relatively close at about twice the median detection capability. At the high a values, the exceedance probabilities are greater for the uniform distribution, as would be expected. In the absence of information from which the real rogue flaw model could be estimated, this figure indicates that the uniform distribution would tend to provide a conservative model if a_{NDE} were greater than 2.

The effective inspection reliabilities for the three rogue flaw models are compared in Figure 62 for the same POD parameters as in Figure 61. In Figure 62, 1-ER(a) is plotted versus a to use the increased resolution provided by the semi-log plot. The ER(a) curves indicate that effective inspection reliability is not rogue flaw model sensitive for the models considered.

The uniform model is a conservative assumption for larger crack sizes when considering exceedance probabilities and effective inspection reliabilities are somewhat insensitive to the three rogue models. Therefore, for the purposes of this study it was assumed that the uniform rogue flaw model would be adequate for evaluating the sensitivity of inspection reliability characterizations. This topic will be further addressed in Section 4.3.

Assuming that rogue flaws will be distributed uniformly over the interval of 0 to A, consideration must be given to a value for A. In the Air Force context, A could conceivably be any number greater than 0.050 in. or 2 to 3 times the median detection capabilities of advanced NDI systems. For some applications, the critical crack length could be quite large. To provide an indication of the effect of A, H(a) and ER(a) were calculated for A=10 and 20 when the median detection capability was 1 and σ=0.5 and 1.0. These results are presented in Figures 63 and 64.

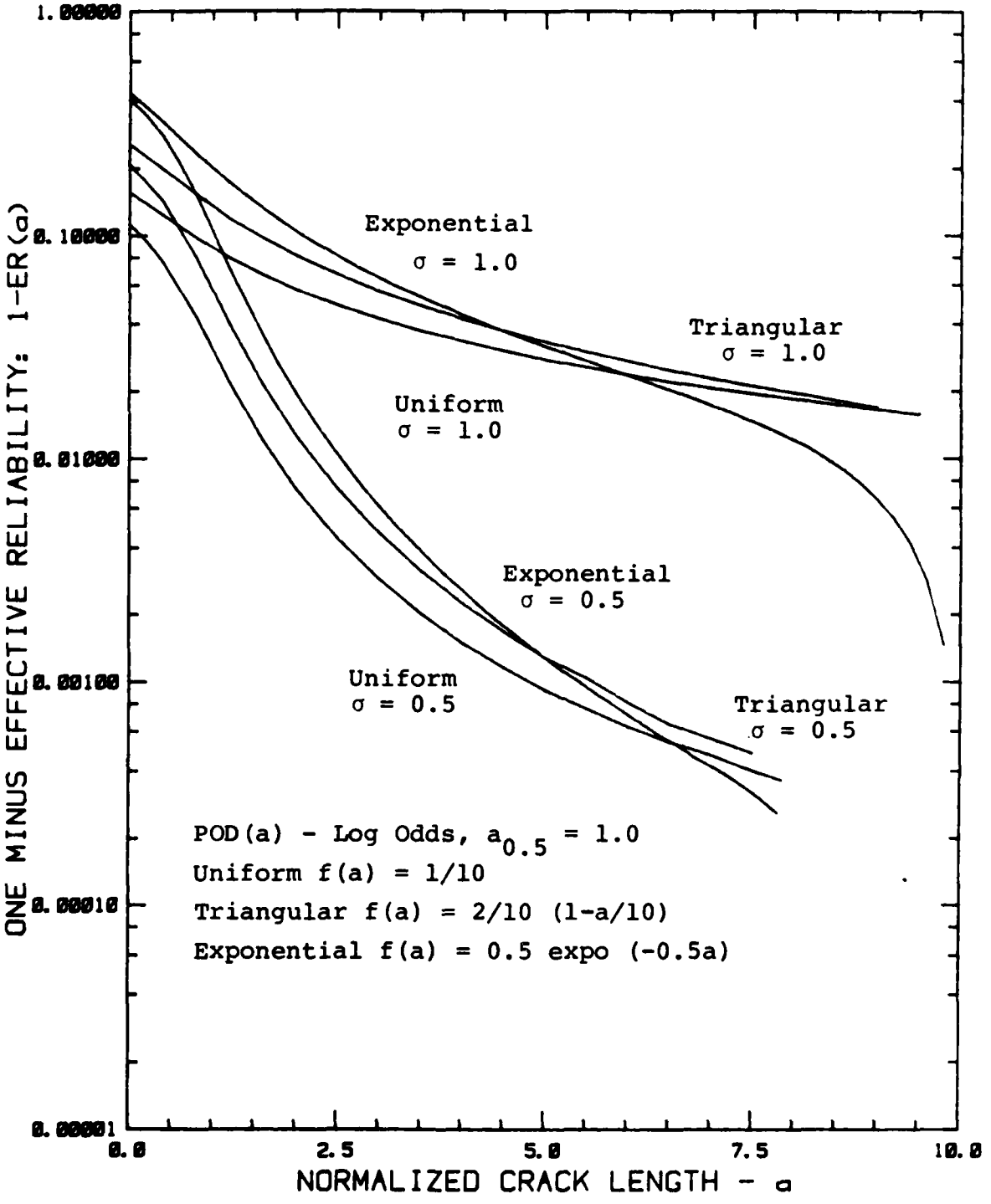


Figure 62 Effective Inspection Reliability for Selected Flaw Crack Size Models - σ = 0.5 and 1.0.

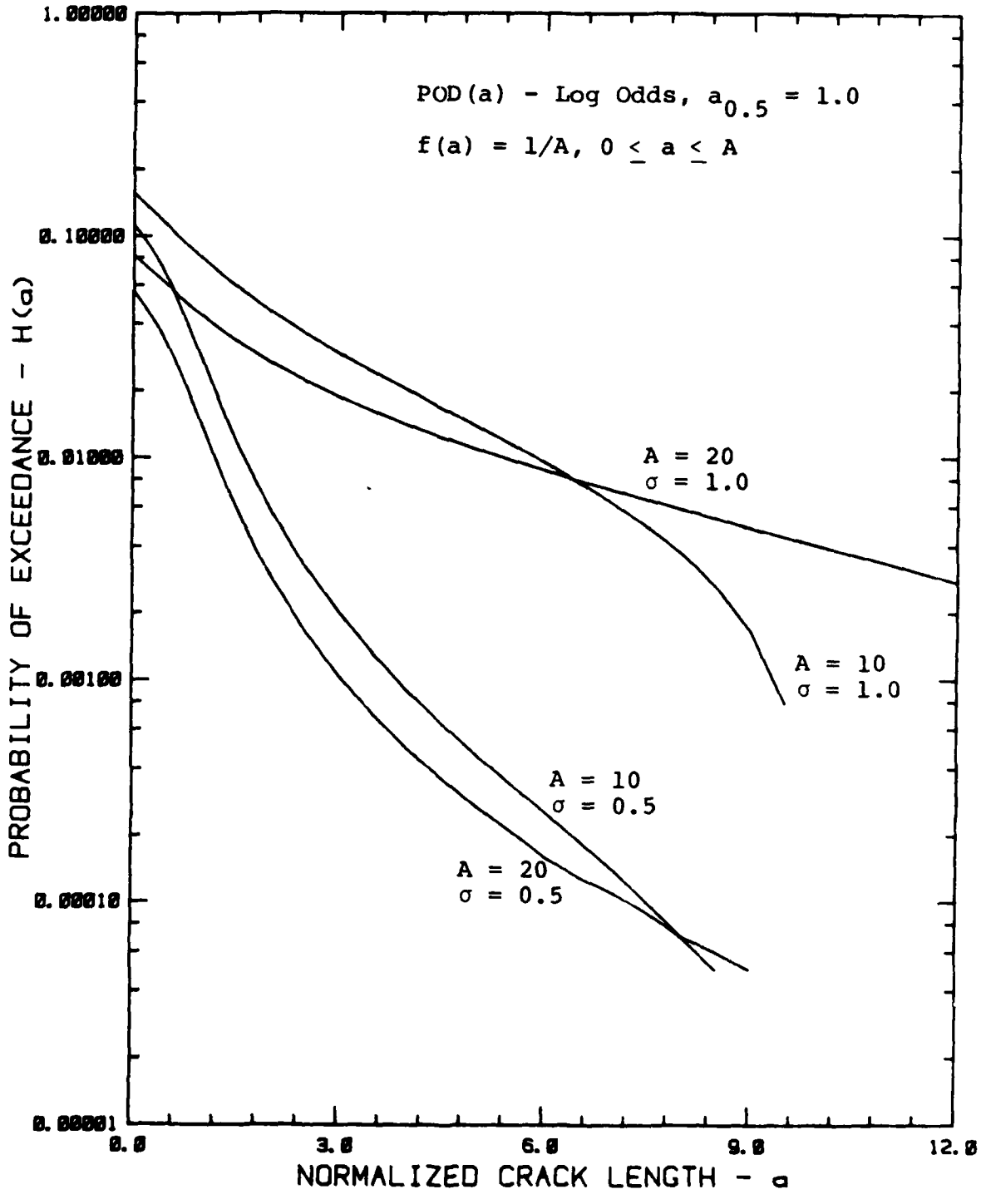


Figure 63. Exceedance Probabilities for Uniform Rogue Flaw Models with 2 Maximum Flaw Sizes - σ = 0.5 and 1.0.

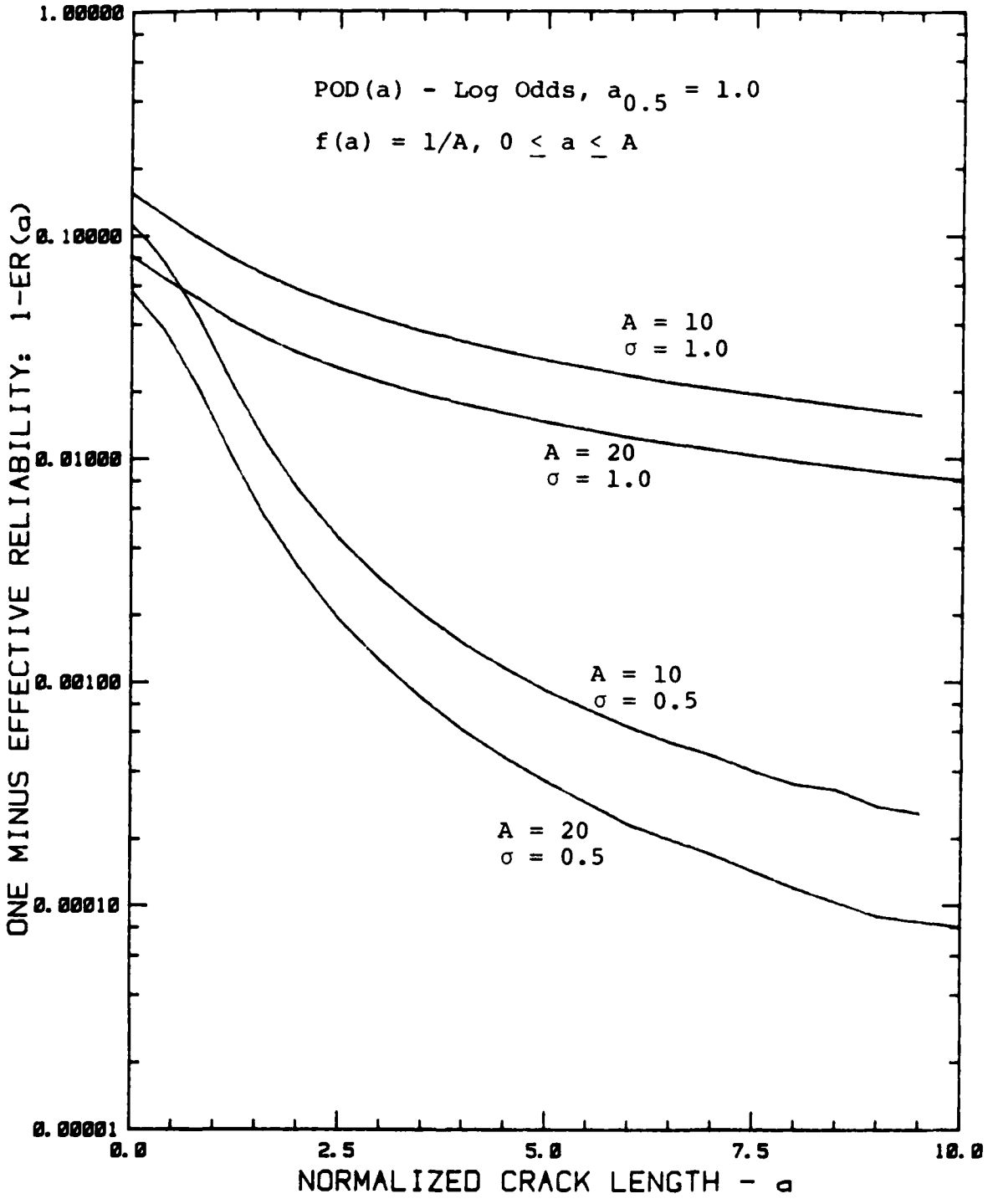


Figure 64. Effective Inspection Reliability for Uniform Rogue Flaw Models with 2 Maximum Flaw Sizes - σ = 0.5 and 1.0.

The effects of POD model are again greater than the effect of A on both the exceedance probabilities and the effective inspection reliabilities. Since A=10 provided larger (more conservative) H(a) values over the practical range of inspection limits and smaller effective inspection reliabilities, this value was chosen for the evaluations of the POD function.

4.2.2 Effect of σ with Uniform Distribution of Rogue Flaws

Figure 65 presents the exceedance probability as a function of normalized crack size for σ=0.1, 0.25, 0.5, 1.0 and 1.5. The effect of σ under these crack size assumptions is quite significant. Note that for the large σ values, an arbitrarily small value of H(a) is not achieved until a is greater than 9. (H(a) must be equal to zero for a $\geq$ 10 since it has been assumed that there are no flaws greater than 10.) Thus, an arbitrarily small exceedance probability can be selected only for inspection systems with small σ (steep POD functions) if the inspection limit is to be of a reasonable size. Figure 66 presents the inspection limits corresponding to exceedance probabilities of 0.01, 0.001, and 0.0001 as a function of σ. For σ less than 0.5, the inspection limits are somewhat insensitive to σ. Again, these exceedance probabilities must be multiplied by the proportion of structural details which contain rogue flaws to obtain the total probability of having cracks greater than the detection limits in the population of details after the inspection.

Since the rogue flaws are much larger than the equivalent crack sizes of the previous paragraphs, the efficacy of the inspection for these flaws is increased. Figure 67 presents the effective inspection reliabilities for the four POD capabilities of Figure 4. This figure indicates that even for σ=1.5 at least 80% of all rogue flaws will be detected and 90% of all rogue flaws greater than twice the median detection capability will be detected. For smaller σ values, the effective inspection reliability is much greater. These results will be further discussed in Section 4.3.

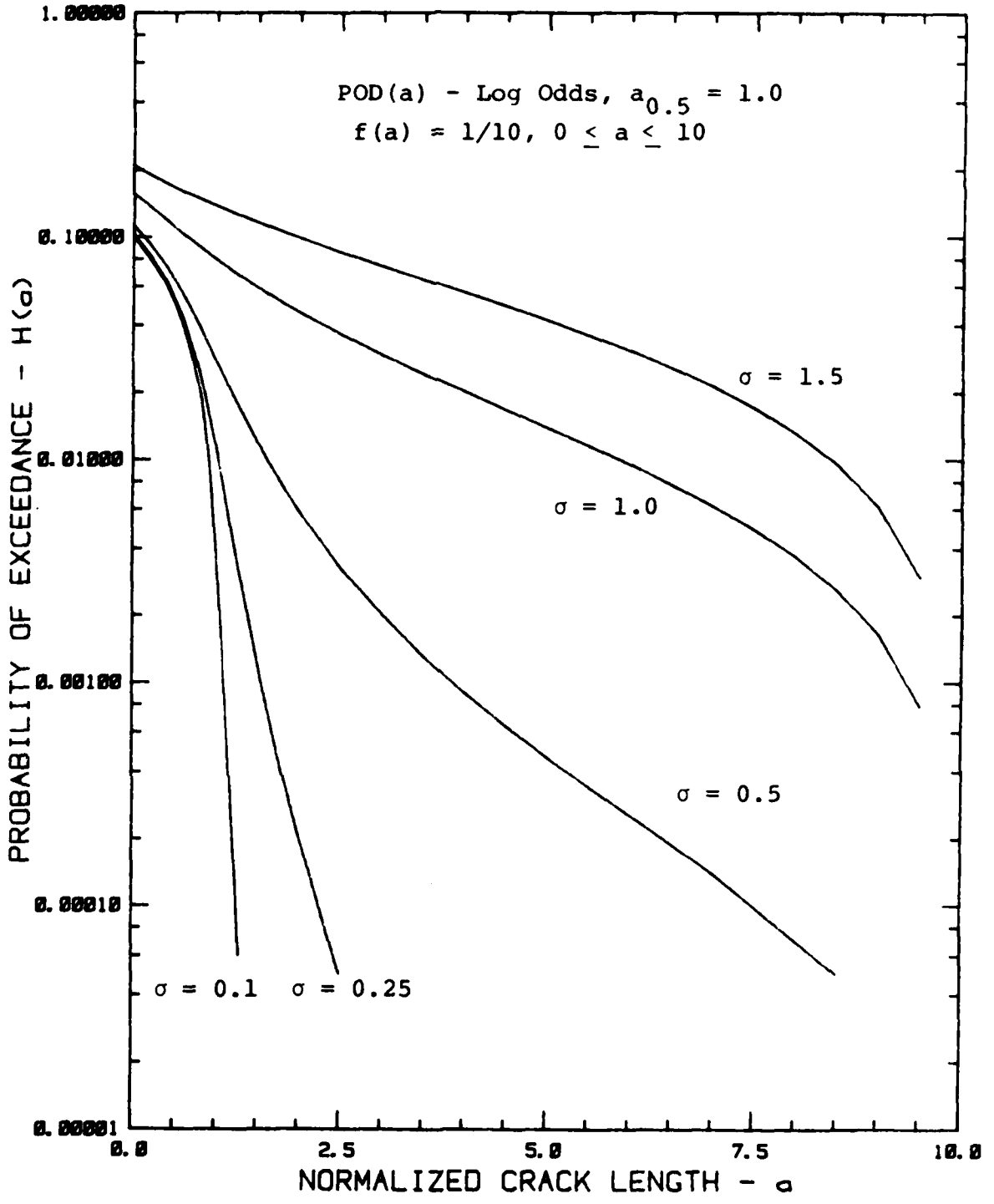


Figure 65. Exceedance Probabilities for Uniform Rogue Flaw Model and Selected σ Values.

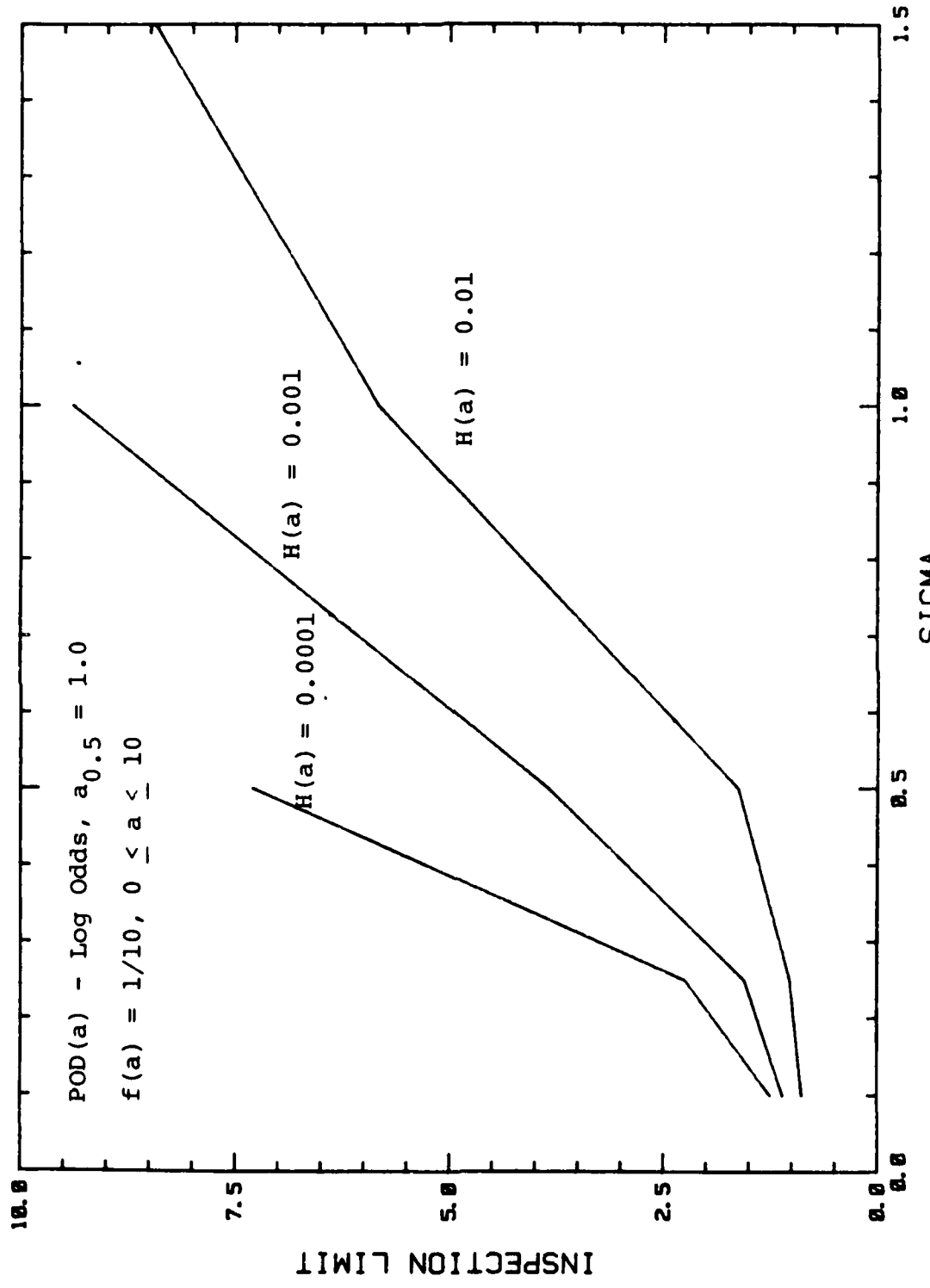


Figure 66. Inspection Limits for Selected H(a) Values as a Function of σ.

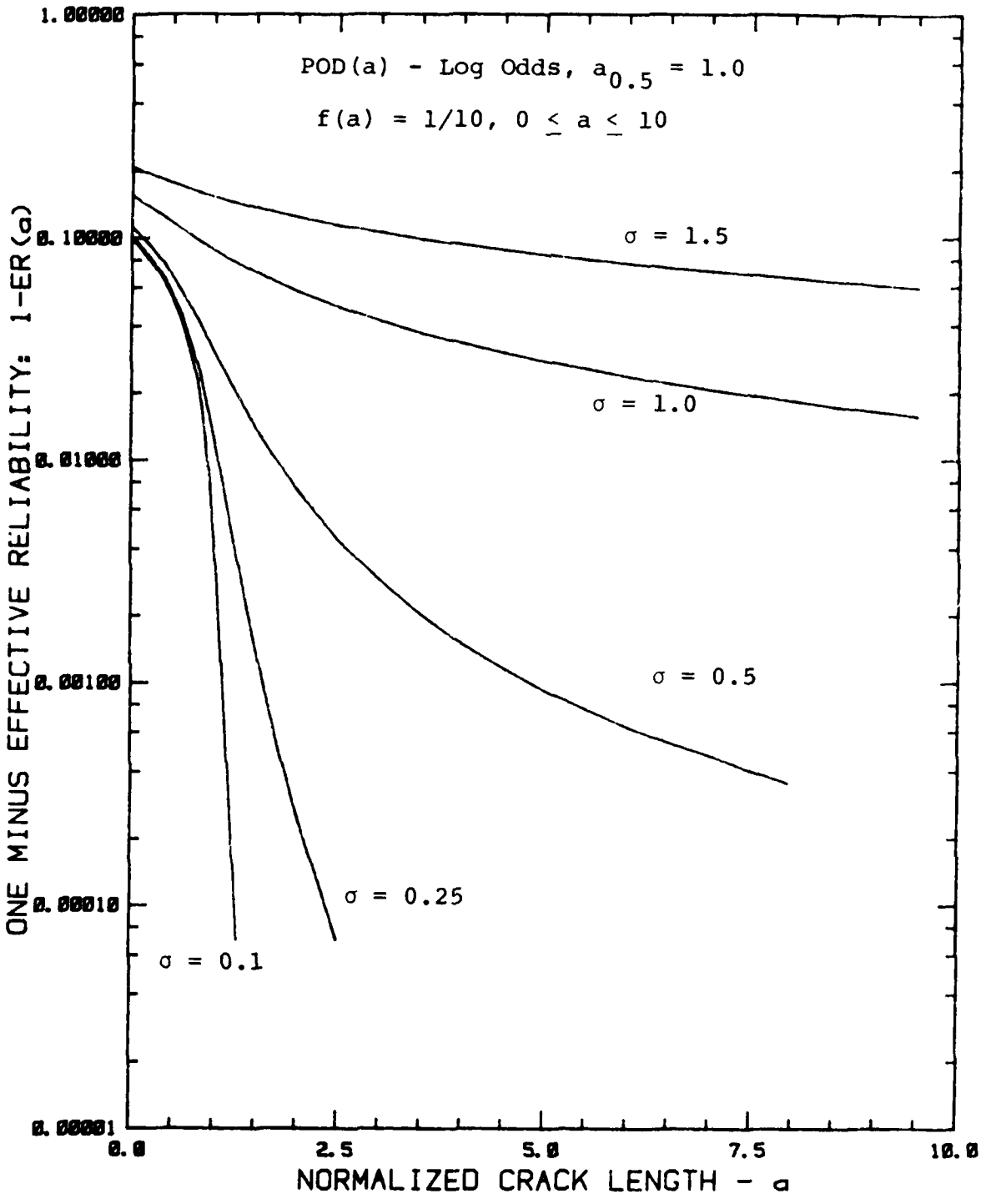


Figure 67. Effective Inspection Reliability for Uniform Rogue Flaw Model and Selected σ Values.

4.2.3 Effect of Median Detection Capability witn Uniform Distribution of Rogue Flaws

To evaluate the effect of median detection capability for fixed σ, exceedance probabilities were calculated for a range of $a_{0.5}$ values with σ=0.5 and σ=1.0. (See Figure 57 for effect of $a_{0.5}$ on POD function itself.) These results are presented in Figures 68 and 69. The range of $a_{0.5}$ values corresponds to the range of $\mu(=\ell n\ a_{0.5})$ of ± 1. Although the pictured range of $a_{0.5}$ values is broad, the effect of $a_{0.5}$ on the calculation of H(a) is significant. This effect can also be seen in the plot of detection limits as a function of median detection capability of Figure 70. This figure (as well as Figure 68) can be used in the evaluation of the statistical variation in the estimation of μ and σ from reliability demonstration programs.

Effective inspection reliability for the five median capabilities with σ=0.5 and 1.0 are presented in Figures 71 and 72, respectively. As anticipated, there is a significant effect of median detection capability on the inspection efficacy. For σ=0.5, increasing the median crack detection capability by a factor of 1.64 results in about a factor of 6 to 7 decrease in effective inspection reliability. For σ=1.0, the effective inspection reliabilities decrease by a factor of 2 to 2.5 for every factor of 1.64 increase.

4.3 DISCUSSION OF CRACK SIZE MODELS

The calculations of the preceeding paragraphs were performed assuming two fundamentally different types of statistical models for the flaw sizes in the population of structural details under consideration. Ideally, a choice between the two assumptions (or the applicability of either) would be made on the basis of real data. However, the availability of such data is extremely limited since, in the absence of an extensive feedback of inspection results and subsequent analysis, the data can only come from teardown inspection. Thus, the choice of the statistical models must be

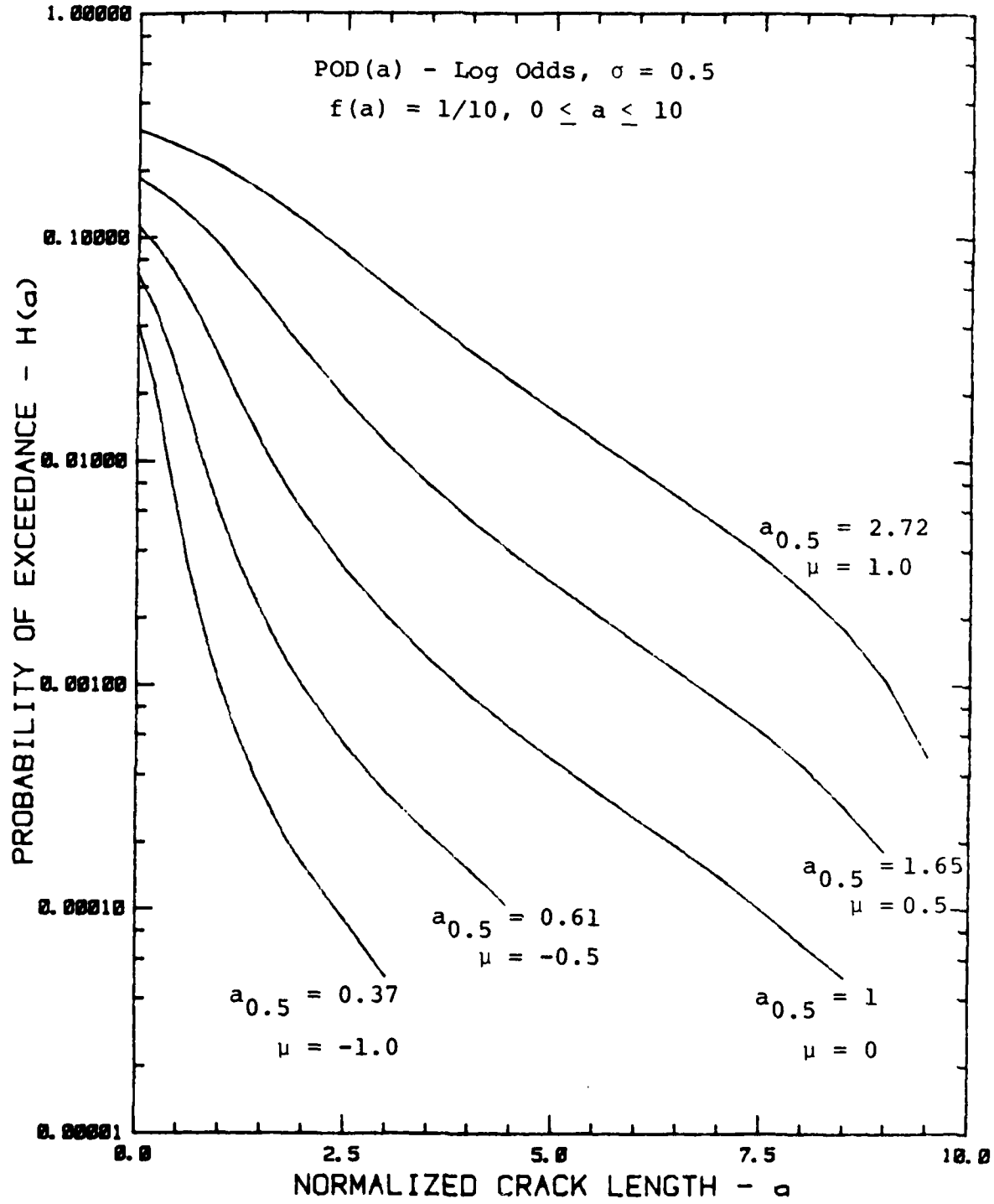


Figure 68. Exceedance Probabilities for Selected Median Detection Capabilities - σ = 0.5.

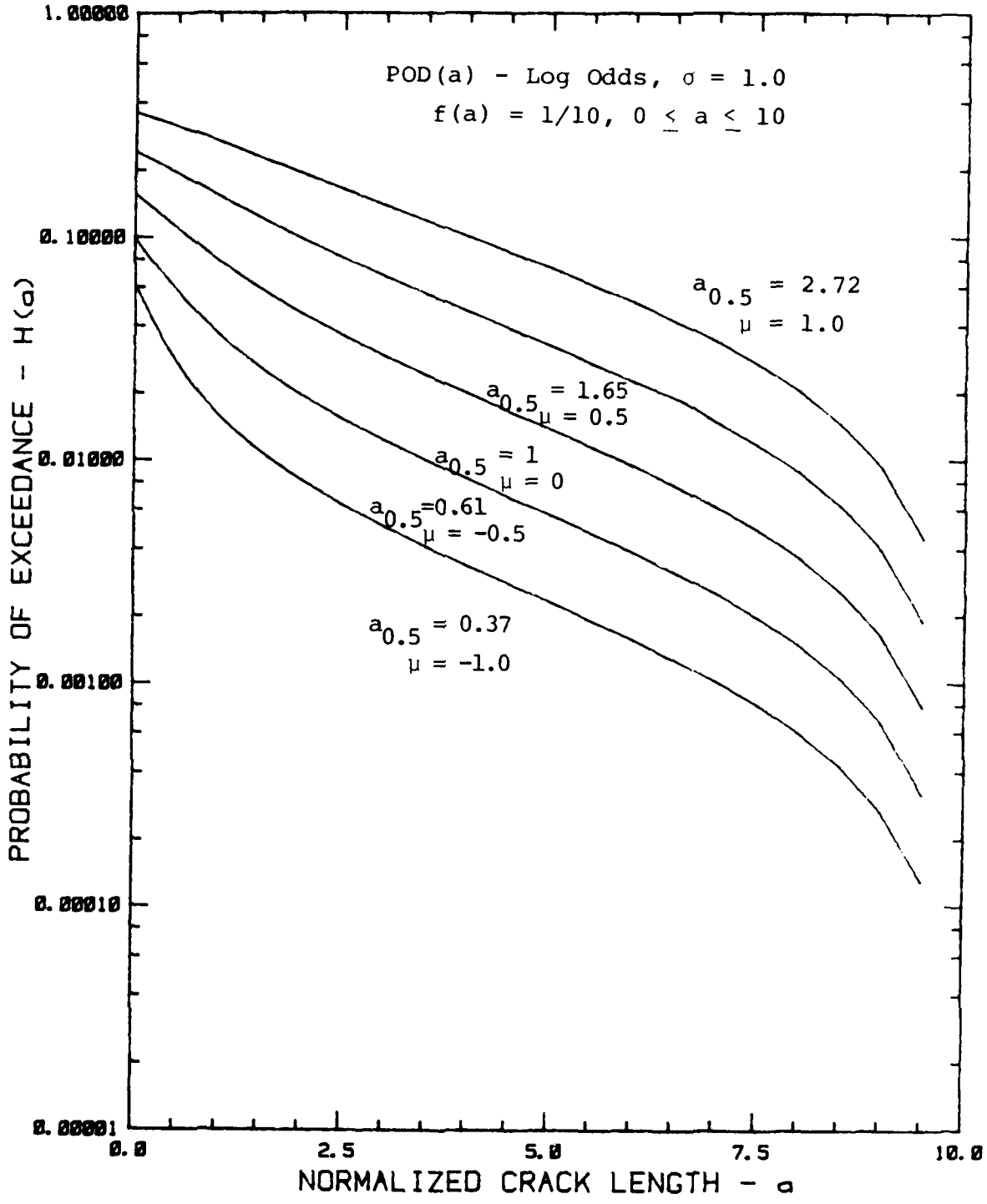


Figure 69. Exceedance Probabilities for Selected Median Detection Capabilities - σ = 1.0

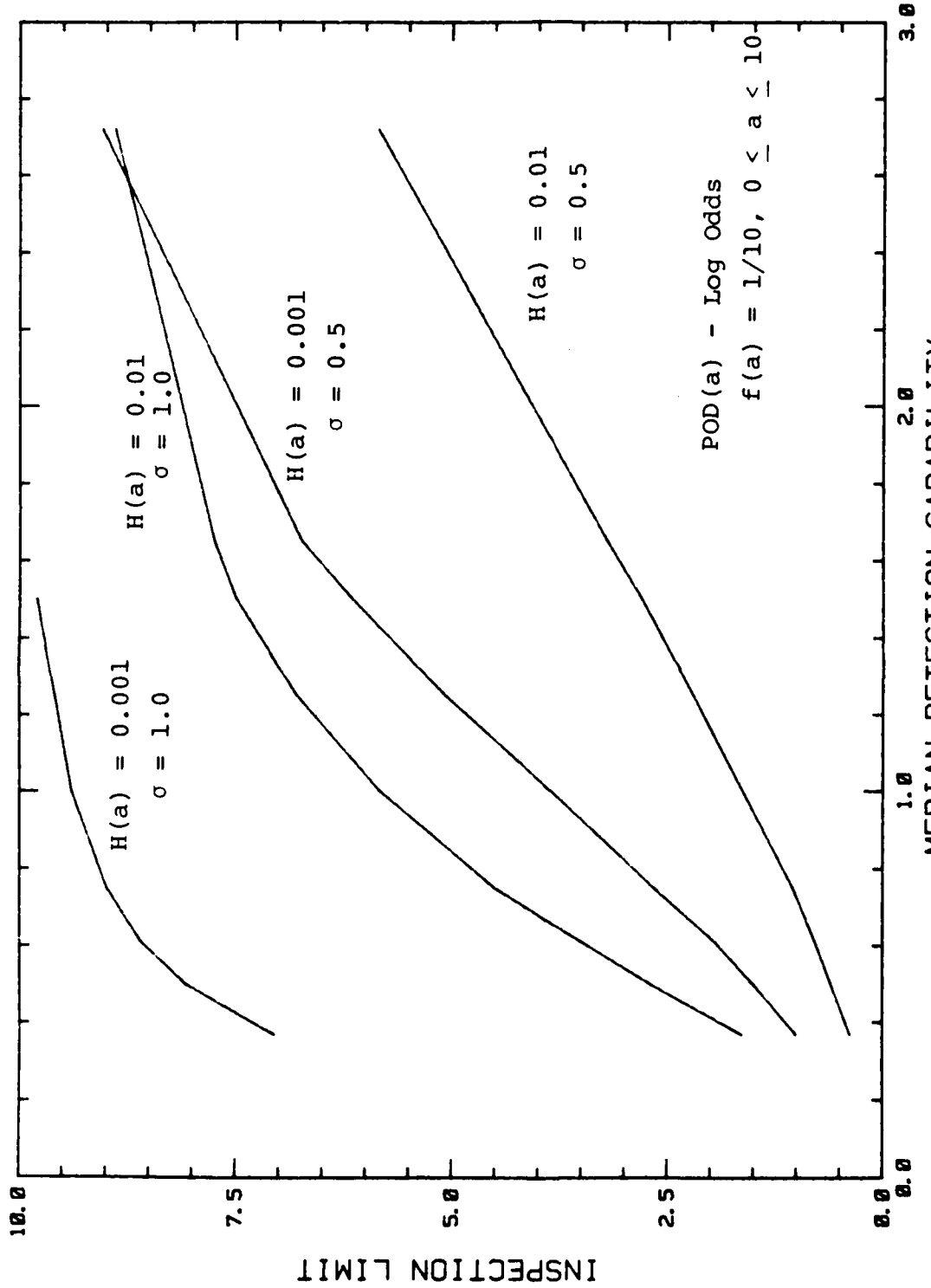


Figure 70. Inspection Limits as a Function of Median Detection Capability for Selected σ and H(a).

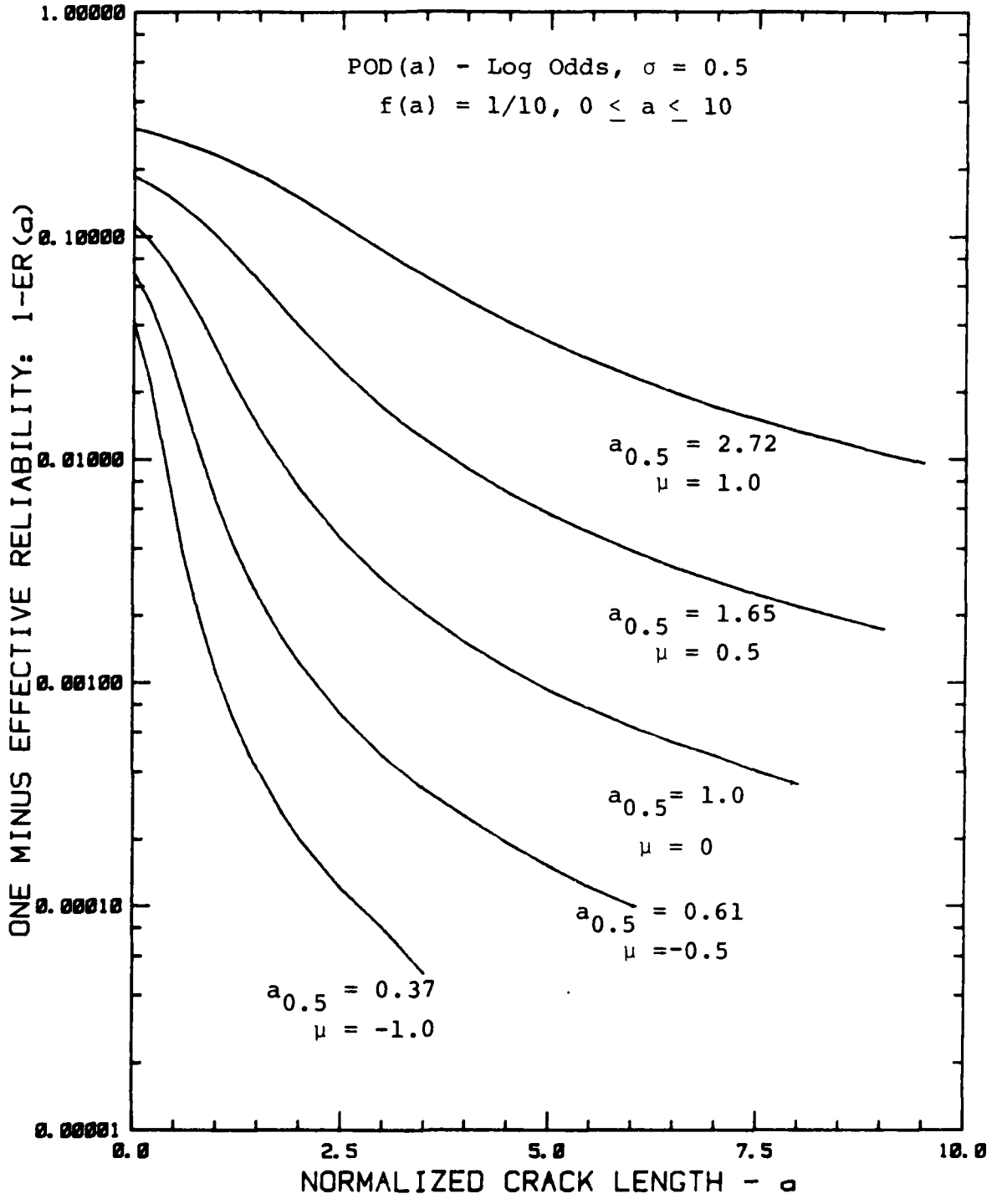


Figure 71. Effective Reliability for Selected Median Detection Capabilities - σ = 0.5.

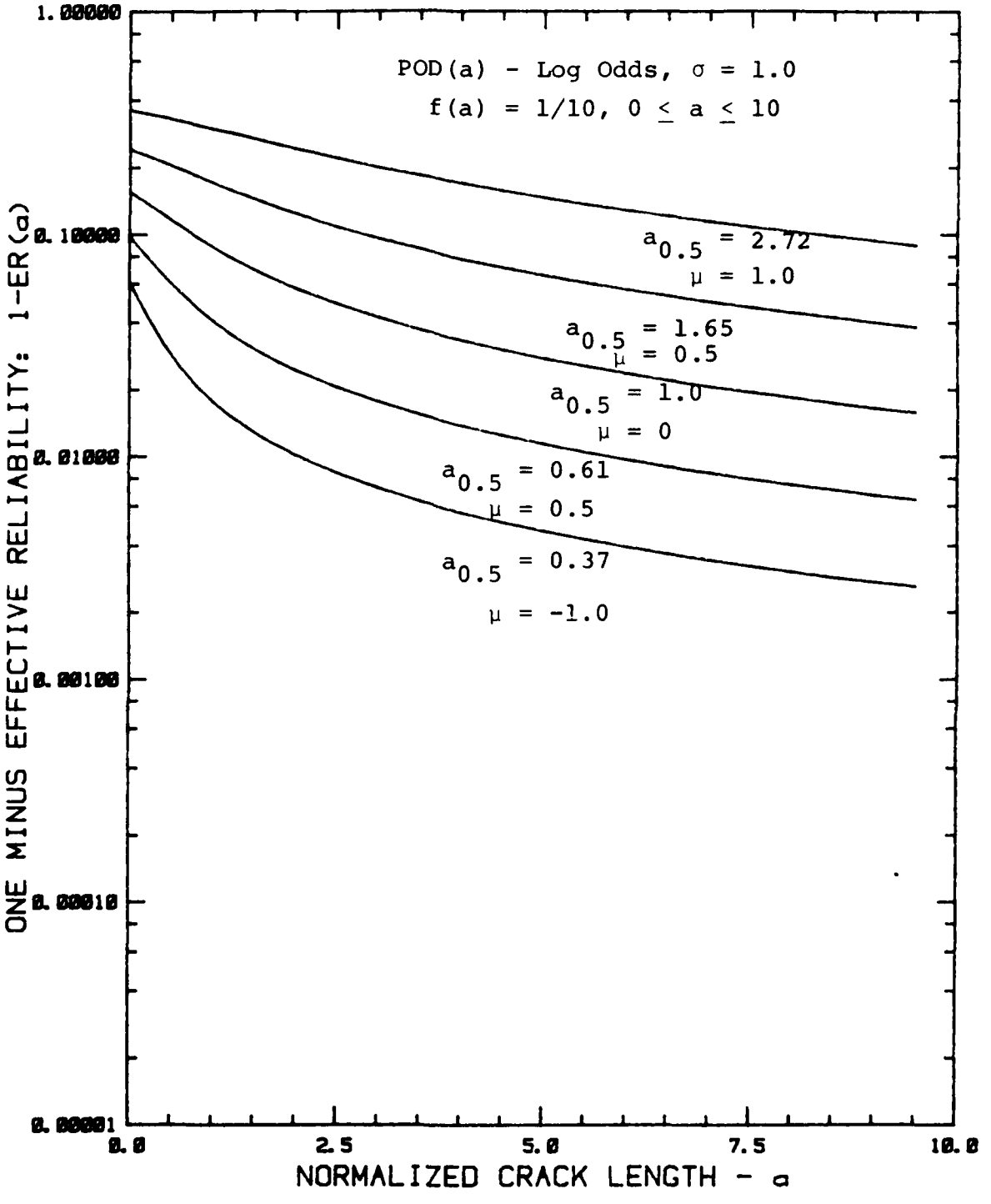


Figure 72. Effective Reliability for Selected Median Detection Capabilities σ = 1.0.

made on an interpretation of the available data and the plausibility of the calculated results.

When a simple form of the crack size density was assumed for the equivalent crack size distribution, the calculation of exceedance probabilities after inspection and the effective inspection reliabilities were somewhat insensitive to the POD function. The crack sizes were generally too small to provide a realistic measure of the effectiveness of the inspection against big cracks. The POD effect on big flaws was masked by the 99.99 percent of details with "no" or extremely small flaw sizes. When the crack sizes were made larger by increasing the median, the results were not realistic in that a high proportion (25 percent or more) of the details were rejected. A higher proportion of large cracks would also result from a density function with a "thicker" upper tail (for example, a log normal distribution of crack sizes) but the expected changes were not considered to be significantly different from those obtained under the Weibull assumption.

While a more complex equivalent flaw size distributional model might reduce the above deficiencies, they are also effectively eliminated by the rogue flaw approach. The rogue flaw model does measure the inspection effectiveness against the larger flaws that may be present in the structure. It also agrees with field inspection data in that few inspections on aircraft structure actually result in a positive call that a defect is present. The major price to be paid for the benefits of this approach is the assumption concerning the proportion of total details which contain "rogue" flaws. This assumption is essentially equivalent to an assumption concerning the right hand tail of the equivalent flaw size distribution. However, field data will give an indication of the total proportion of flawed specimens whereas equivalent flaw size models tend to be dominated by the extremely small crack sizes.

For the above reasons, the rogue flaw model was selected for the characterization of inspection capability for damage tolerance analyses. Further, it will be assumed that the rogue flaws will be uniformly distributed between zero and an upper limit (nominally considered to be the critical crack size). The uniform distribution is considered conservative and is quite often assumed by statisticians in the absence of information to the contrary. The use of other distributions (e.g. the exponential and triangular) in the rogue flaw model were judged to have an insignificant impact on the conclusions of this study.

SECTION 5
ASIP APPLICATIONS

The United States Air Force Aircraft Structural Integrity Program (ASIP) is mandated by Air Force Regulation 80-13 and defined by the requirements of MIL-STD-1530A. The damage tolerance design requirements are specified in MIL-A-83444 and are imposed on the airframe manufacturers to insure safety and operational readiness throughout the life of the aircraft. These requirements force the manufacturer to demonstrate through analyses and tests that the airframe will not fail during periods of unrepaired usage even if critical structural elements contain damage or flaws of a pre-specified size. If necessary, the manufacturer is permitted to assume depot or base level inspections at intervals of about one-fourth of the design lifetime.

The MIL-A-83444 requirements are design requirements and are based on assumptions concerning design usage and stress levels. Ideally, no base or depot level damage tolerance inspections would be required in the design life of the aircraft (a goal achieved by the F-16). However, if mid-life inspections were necessary, presumably, they would be scheduled in accordance with the MIL-A-83444 analyses and implemented through the Force Structural Maintenance Plan which is required by the Force Management Tasks of MIL-STD-1530. These inspections could also be called for due to the update of analyses based on actual operational usage or the occurrence of unanticipated problems.

Of prime concern to the objectives of this study are the MIL-A-83444 initial flaw size and in-service inspection flaw size assumptions. These flaw size assumptions are quite specific but it is generally assumed that the specified sizes are subject to change and the amount of change will depend on NDI capabilities. Since the initial flaw size assumption is considered to be distinct from the in-service inspection flaw size assumption, these topics will be addressed separately in the following.

5.1 INITIAL FLAW SIZE ASSUMPTION

Initial damage assumptions are applicable to two categories of structure (slow crack growth and fail safe); for maximum damage in the most critical hole or cutout and a non-hole location of each structural element; and for every hole in every detail to represent manufacturing quality for continuing damage calculations. Without loss of generality, the following discussion will focus on the 0.050 in. crack assumed to be present at the most critical hole or cutout in slow crack growth or non-inspectable structure. The corresponding crack size assumption in fail safe, multiple load path structure is 0.020" in MIL-A-83444. The initial crack at a non-hole or cutout location is generally not the driver of structural life. The 0.005 in. crack which is assumed to be present in every hole as a measure of manufacturing and processing operation is far below the reliable detection capability of current NDI systems.

When MIL-A-83444 was first issued the initial flaw size assumptions for the critical location (e.g. the 0.050 in flaw) were assumed to be indicative of inspection capability for a 0.9 POD at either a 95% confidence level for slow crack growth structure or a 50% confidence level for fail safe structure [15]. MIL-A-83444 does contain a provision for lowering the initial crack sizes in slow crack growth structure if an approved NDI reliability demonstration program yields a 90/95 crack size limit less than the specified values and the NDI system is applied during quality control to all the relevant critical structures.

A broader interpretation of the initial crack sizes is now being made in that the initial crack size assumption is also considered to cover manufacturing defects which are not necessarily amenable to detection by NDI systems. Examples of such defects would be out-of-round or mislocated holes. The equivalence between such defects and a crack size is based on engineering judgement and experience (satisfaction) with the current initial crack size assumptions.

Regardless of the question of the initial crack sizes being equivalent to non-detectable manufacturing or processing flaws, the provision for lowering the initial crack size assumption based only on a lower 90/95 detectable crack length should be removed from the specification. This recommendation is based on two reasons.

First, the 90/95 crack length characterizes the NDI reliability on the basis of a single point of the POD function. This one number characterization of inspection capability does not permit risk evaluations or cost benefit analyses since it is decoupled from the structure [16]. Further, use of this characterization could lead to increased failure risks as shown by the following example. Figure 73 displays a hypothetical distribution of flaw sizes and two POD functions representing two inspection reliabilities. Given NDI reliability programs with very large sample sizes, the 90/95 limits for $P_1(a)$ would approach the "true" value of 0.025 while that of $P_2(a)$ would approach 0.050 as indicated by the arrows. Under the current specification, the initial crack size assumption could be lowered to 25 mils. Figure 74 displays the exceedance probabilities for the two inspection capabilities and the assumed crack size density. Under the 50 mil initial crack size assumption, there is a probability of 0.000035 of having a crack greater than a_o (=0.050 in.) in the structure after the inspection. Under the increased inspection reliability (i.e. the 90/95 crack size = 25 mils) there is a probability of 0.005 of having a crack greater than a_o(=0.025 in.) in the structure after the inspection. To maintain the same exceedance probability as for the original inspection, as characterized by $P_2(a)$, a_o should only be reduced to 0.046 in. Therefore, reducing a_o to 90/95 crack lengths may increase the probability of having cracks greater than a_o in the structure and, hence, increase the risk of failure.

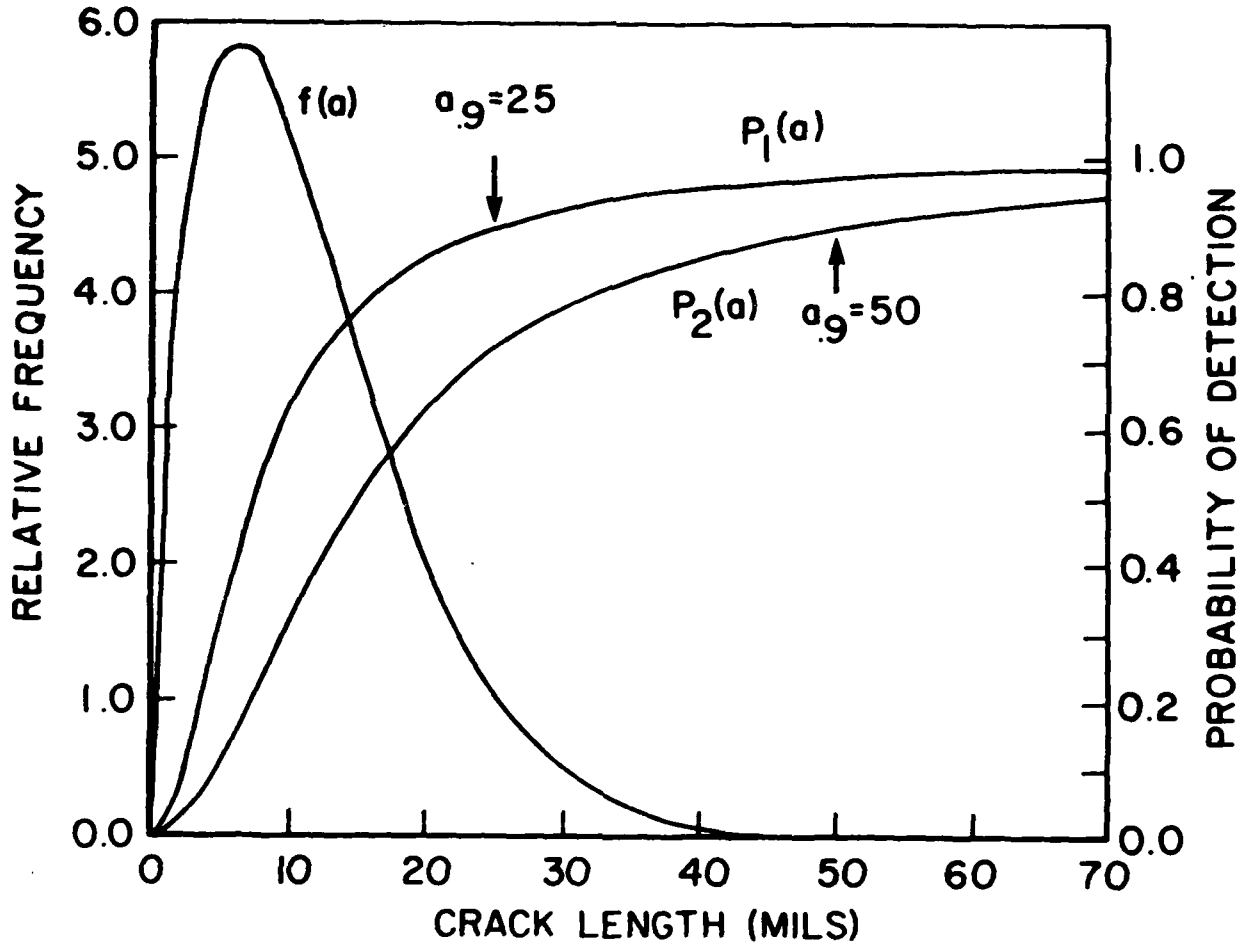


Figure 73. Example POD Functions for Increased NDI Capability.

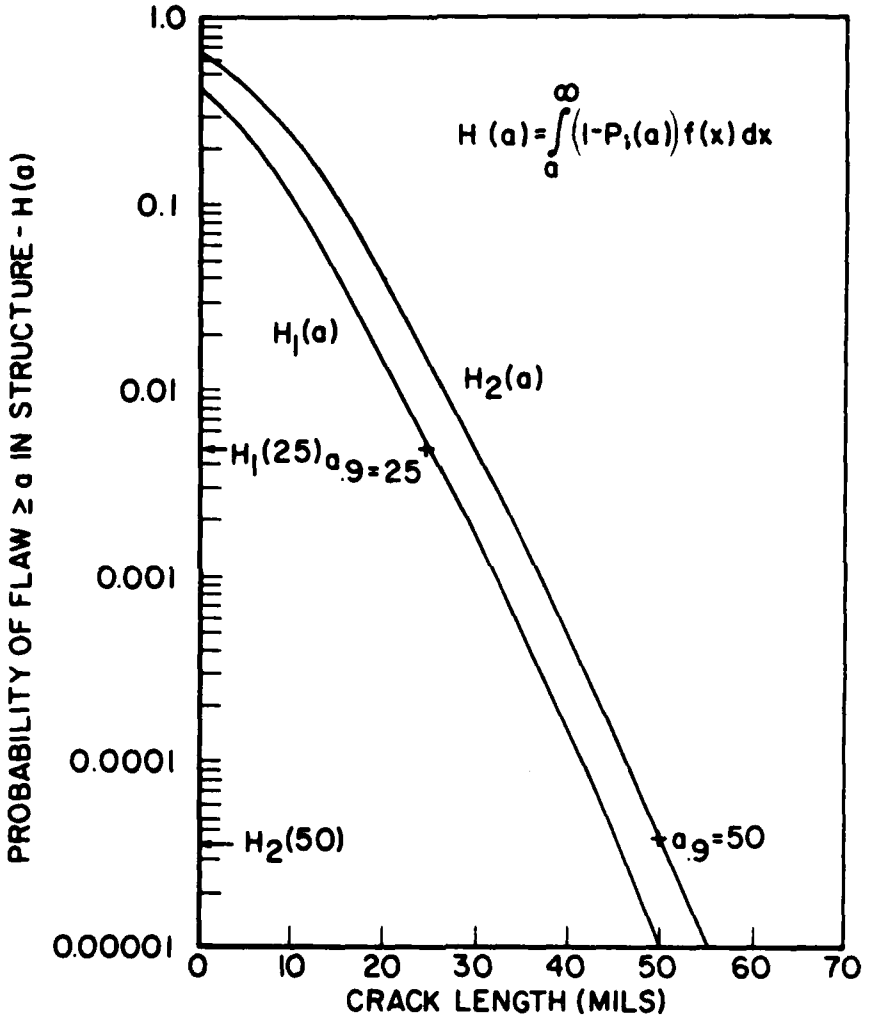


Figure 74.. Exceedance Probabilities for Increased NDI Capability.

Different assumptions would, of course, result in different exceedance probabilities but the basic conclusion would remain the same.

The second reason for not using 90/95 crack length characterizations of NDI reliability pertains to the unacceptably large degree of scatter in the estimates of POD/CL type limits at the high values of POD and confidence required to insure structural integrity. In a previous study [2], NDI reliability experiments were simulated and selected POD/CL limits were calculated for each simulated experiment based on the "inspection" of 400 cracks. The entire "experiment" was repeated 100 times under a known POD function. Figure 75 presents histograms of the 90/95 and 95/90 limits for the simulation of 100 reliability experiments with the known POD function as shown in the figure. Estimating a 90/95 limit by an NDI reliability experiment with this capability is approximated by drawing a value at random from the cross-hatched histogram. For this POD function, the crack length for which POD(a) = 0.9 is 20.1 mm. the average of the 90/95 limits were 45.0 mm with a total range of 25 to 85 mm. These example results illustrate that a 90/95 limit is not a precisely defined characteristic of NDI reliability but rather is a random quantity with a relatively large degree of scatter. Thus, consecutive NDI reliability experiments on the same equipment using different cracked specimens could produce significantly different 90/95 values and neither of them need be close to the true crack length for which the POD is equal to 0.9. The POD/CL limits become even less stable as the POD or the confidence level increase.

The above paragraphs demonstrate that the 90/95 type NDI reliability characterization should not be used as a basis for lowering the crack size assumptions for damage tolerance analysis. In fact, due to the extreme scatter in these estimates, the 90/95 characterization should possibly not be used to compare the reliability of different NDI systems. Considerable caution is required in the interpretation of 90/95 limits once they are removed from the complete POD function.

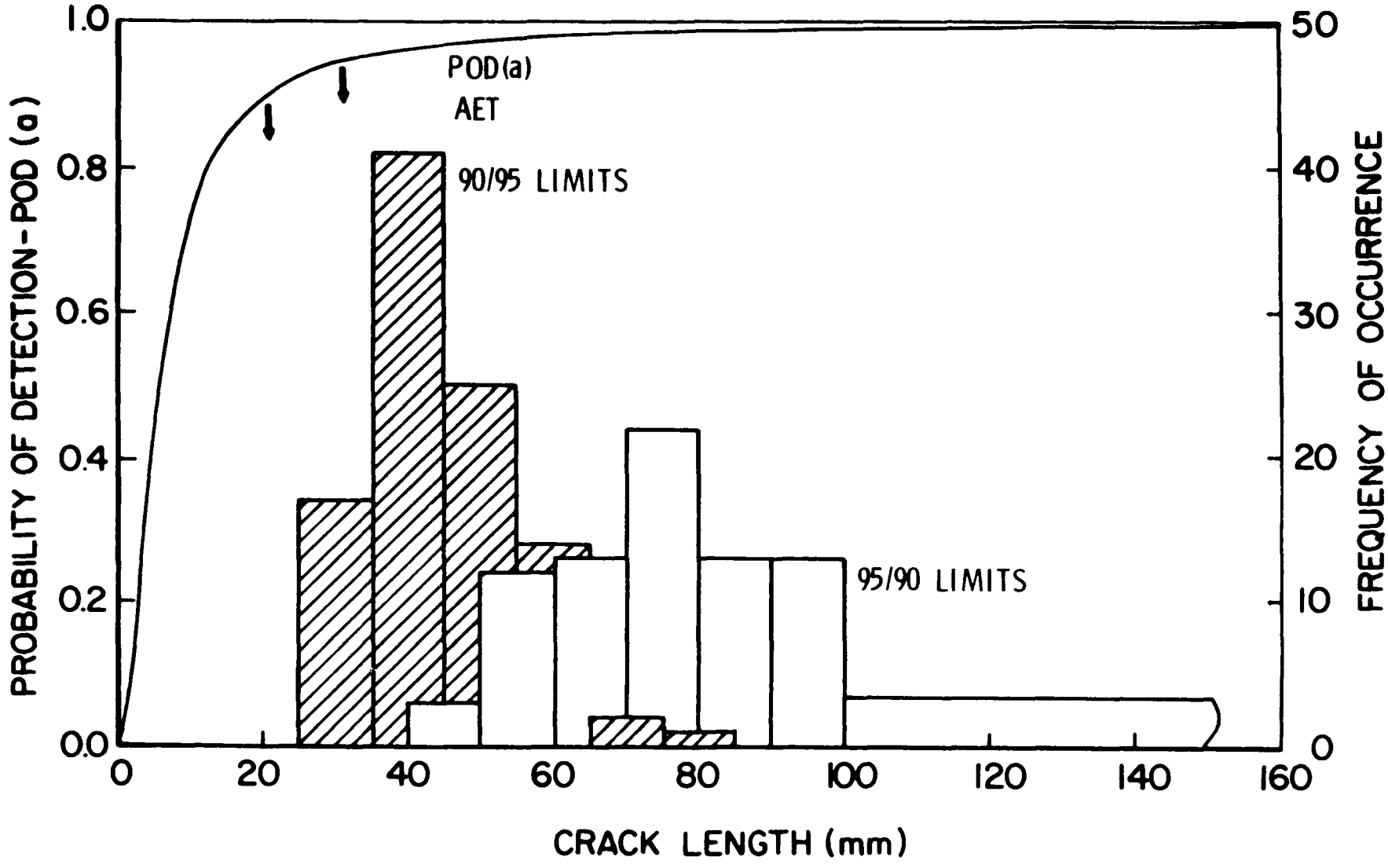


Figure 75. Example Distribution of 90/95 and 95/90 Limits from Simulated NDI Reliability Experiments.

Even though the initial crack size assumptions of MIL-A-83444 are not expected to be lowered based on inspection capability alone, it may be desireable to evaluate the 0.050 in. assumption or any other in terms of its likelihood of being exceeded. It is recommended that this evaluation be expressed in terms of the exceedance probability and/or the effective inspection reliability evaluated at a_o. The recommended assumptions and methods for making these calculations will be discussed in the next subsection as they may be more germane to the evaluation of in-service inspections.

5.2 IN-SERVICE INSPECTION FLAW SIZE ASSUMPTIONS

The MIL-A-83444 post in-service inspection flaw sizes were based on a broad brush evaluation of the limited amount of data then available from which depot or base level inspection reliability could be inferred [14]. When more data did become available [1], they indicated that depot level inspection reliability as quantified by a 90/95 crack length characterization was, in general, poorer than that assumed in MIL-A-83444. The current depot capability is still considered at least somewhat unknown [16] but it is not clear what is meant by a general depot level capability.

MIL-A-83444 is considered by the Aeronautical Systems Division to be a design specification. The flaw size assumptions of this document are used by the manufacturer to demonstrate that the airplane's structure will be damage tolerant. To date, in-service inspection requirements to meet damage tolerance objectives have only been performed as a result of damage tolerance assessments on pre MIL-A-83444 aircraft or in response to specific problems. In these applications, post inspection flaw sizes have been decided on a case-by-case basis taking into account all features of the structural element that could affect the detectability of the flaw being sought. Such features include the specificity and accessability of the problem area; the material, geometry, and physical condition of the inspected detail; the consequences of failure versus cost of inspections;

and a characterization of the proposed NDI system's reliability. In general, there is very little or no NDI reliability data for this detailed level of application. Thus, all of the features are subjectively evaluated and a consensus based on engineering judgement is used to arrive at the reset crack length for prediction of future maintenance or retirement actions.

In view of the above there are two types of post-inspection flaw size assumptions: the general measure of depot or base level reliability (e.g. as stated in MIL-A-83444) and the component specific, reset crack length derived for each real application. To distinguish between the two, define the inspection limit (a_{NDE}) to be the general measure of NDI reliability and the reset value (a_{RS}) to be the component specific measure. Then, an inspection limit is one of the factors which enter into the establishment of the reset value. This study has been directed at methods of characterizing and estimating inspection limits.

Modifications to the post-inspection flaw size assumptions of MIL-A-83444 may be unwarranted since very little experience has been realized with the current inspection limits and there is no hard evidence to demonstrate that the overall Air Force depot level capability is better or worse than that indicated. However, two recommendations can be made. As more emphasis is placed on NDI reliability, POD functions will be generated for many NDI system/application combinations. These will lead to classes of generic capabilities and a refined definition of Air Force inspection limits will be possible. That is, one crack length (e.g. 0.25 in.) may not be a sufficient description of NDI reliability for all aircraft structure. Rather, different categories of inspection limits will be possible as defined by type of material, access to automated equipment, coatings, etc.

The second recommendation concerns the method of characterizing NDI capability in terms of a single crack length. It is assumed that the current characterization is based on a POD/CL crack length type characterization and, for the reasons discussed in Subsection 5.2, it is recommended that this characterization

not be used in the future. To evaluate inspection reliability for crack size assumptions, it is recommended that NDI systems be characterized in terms of the exceedance probabilities, H(a), as calculated from equation (1) or the effective inspection reliability, ER(a), as calculated from equation (7). Pending further information on the distribution of flaw sizes that may exist in the structure, a uniform distribution of crack sizes under the rogue flaw model approach is recommended. For a specific structural detail, the flaws sizes should be uniformly distributed between zero and the critical crack size. For a generic characterization, the flaw sizes should be distributed between zero and a universally agreed upon multiple (say 10 or 20) of the median crack detection capability.

Given the H(a) and ER(a) functions, crack sizes can easily be determined which will yield a fixed probability of exceedance or a fixed effective inspection reliability. Experience will be required with these measures to see if one will be preferred. If a single POD function model (such as the log odds model) becomes generally accepted, these evaluations could be made using one set of normalized curves as, for example, in Figures 65 and 67. Interpolating to the correct σ value and denormalizing with the median detection capability would be the only required calculations.

As an example, assume that a particular NDI system application has a log odds model with $\mu = 0.020$ in. and $\sigma = 0.5$. If the inspection limit is to be determined such that one rogue flaw in 1,000 is to pass undetected, then from Figure 65 or Figure 66 the normalized detection limit would be 3.8 and the denormalized detection limit would be about 3.8 (0.020) = 0.076 inches. Similarly, if the desired effective inspection reliability is to be 0.999, then, from Figure 67, the normalized characterizing crack length would be 4.8 which corresponds to 4.8 (0.020)= 0.096 inches. Thus, with this inspection system (with rogue flaws distributed uniformly between 0 and 0.2 in.), there is a probability of 0.001 that a rogue greater than 0.076 inches would be in the structure and not detected. Also, 99.9 percent of all

flaws greater than 0.096 inches will be detected. Recall that only a small proportion of details have rogue flaws so that the proportion of all details which will contain undetected flaws is considerably smaller than 0.001.

Finally, it is noted that to reflect the uncertainty associated with the estimate of the POD function, a lower bound on the entire POD function can be used in the calculations of equations (1) and (7). Paragraph 3.3 provides methods for calculating the lower confidence bound function.

SECTION 6

CONCLUSIONS

This program comprised three somewhat distinct phases. The major conclusions drawn from each of the phases are summarized in the following.

6.1 POD ESTIMATION

Probability of detection as a function of crack size is the basic element in characterizing inspection uncertainty. This study assumed that POD(a) is adequately modeled by the log odds function. Methods of estimating the parameters and placing confidence limits on the log odds POD function were developed and applied to simulated NDI expeirments for different types of data and analyses procedures. The conclusions were as follows.

1. When inspection results are only available in a pass/fail format, maximum likelihood estimation procedures should be used if there is one inspection per flaw. If multiple inspections per flaw are planned, at least 10 inspections should be performed for each flaw and the regression analysis approach should be used. The extra binomial variation model (i.e. assuming a beta distribution for individual crack detection probabilities) requires as many or more inspections per flaw with a large increase in complexity. Further, confidence bounds on the POD function have not yet been developed using this approach.

2. If the NDI system response signal, $\hat{a}$, is available, the $\hat{a}$ vs a approach to estimating the POD function should be used. This analysis is approximately twice as efficient as that of the pass/fail method in a well designed experiment. That is, the pass/fail method requires twice as many samples to obtain the same estimation precision as that of the $\hat{a}$ vs a approach.

The equation for estimating the parameters and placing confidence bounds on the POD function are presented. This data format also provides a convenient approach to testing the log odds model or to developing different models. The $\hat{a}$ vs a analysis method reduces the effect of a poor crack size selection in a demonstration program and also permits an analysis seeking the causes of variation under designed experiments with multiple observations per flaw.

3. In designing NDI reliability demonstration experiments it is necessary to insure that the flaw sizes in the sample specimens span the range of the POD function. Ideally, the cumulative distribution of the sample specimen sizes would be approximately the same as the POD function. The number of sample specimens should be at least 30 if they have the appropriate size distribution and more, if not.

6.2 NDI RELIABILITY CHARACTERIZATION

While the POD function quantifies inspection uncertainty, current inspection capabilities cannot guarantee that virtually all cracks greater than a practical size for damage tolerance analyses will be found at an inspection. Therefore, an NDI reliability characterization was derived which correlates inspection uncertainty with the probability of having "large" flaws in a structure after inspection. The following conclusions regarding this characterization were drawn.

1. Any link between inspection uncertainty and risk of structural failure must account for the sizes of flaws in the population of details before the inspection. The simplest links are the exceedance probabilities and the effective inspection reliability which are calculated from the POD function and the pre-inspection flaw size distribution.

2. Meaningful inspection crack size limits for reflecting NDI reliability in particular applications can be calculated from the exceedance probability function, H(a), or the effective inspection reliability function, ER(a). From the exceedance

probability function, an inspection limit can be determined such that there is a pre-specified probability (presumably very small) of having a crack greater than the inspection limit and missing it. From the effective inspection reliability function, an inspection limit can be determined such that a predetermined proportion (presumably very large) of all cracks greater than the inspection limit will be detected. A choice between these criteria has not been made pending experience with applications.

3. Very little information is available concerning the distribution of flaw sizes in aircraft structure. An equivalent initial flaw size model with attendant fatigue growth did not prove viable. A rogue flaw approach wherein only actual flaws in the structure are included, was formulated. Assuming the rogue flaws have a uniform distribution between 0 and the critical crack size provided generally conservative inspection limits.

6.3 ASIP APPLICATION

The conclusions regarding the use of the flaw detection reliability characterization in ASIP applications are as follows.

1. The initial crack size assumptions for as-manufactured structural details are intended to cover manufacturing defects which are not necessarily detectable by conventional NDI systems. The provision for lowering the initial crack size based on a 90/95 type inspection reliability characterization should be removed from MIL-A-83444.

2. The in-service inspection flaw size assumptions of MIL-A-83444 are values for computation during design and development stages of an aircraft's life and are intended to be general measures of depot and base level inspection capability. These crack sizes should be re-evaluated based on generic estimates of POD capability and the exceedance probability function, H(a) and the effective inspection reliability function ER(a).

3. For determination of reset crack length after in-service inspections of specific structural detail, inspection capability is one of several factors which are evaluated. The best estimate of the POD function for the specific application should be used with the uniform rogue flaw model between 0 and the critical crack length to calculate inspection limits as defined by H(a) and/or ER(a).

4. Experience is needed with respect to the evaluation of POD function and their characterization in terms of the H(a) and ER(a) functions.

REFERENCES

1. Lewis, W. H., et. al., "Reliability of Nondestructive Inspections - Final Report," Report No. SA-ALC/MME 76-6-38-1, United States Air Force, San Antonio Air Logistics Center, Kelly Air Force Base, Texas, 1978.

2. Berens, A. P. and P. W. Hovey, "Evaluation of NDE Reliability Characterization," AFWAL-TR-81-4160, Air Force Wright Aeronautical Laboratories, Wright-Patterson Air Force Base, Ohio 45433.

3. Gallagher, J. P., et al., "USAF Damage Tolerant Design Handbook: Guidelines for the Analysis and Design of Damage Tolerant Aircraft Structures," AFWAL-TR-82-3073, Air Force Wright Aeronautical Laboratories, Wright-Patterson Air Force Base, Ohio 45433.

4. Tiffany, C. F., "The Role of Inspections in the Maintenance of Aircraft Safety," Proceedings of the Government/Industry Workshop on the Reliability of Non-Destructive Inspections, Report No. SA-ALC/MME 76-6-38-2, United States Air Force, San Antonio Air Logistics Center, Kelly Air Force Base, Texas, August, 1978.

5. Yee, B. G. W., et al., "Assessment of NDE Reliability Data," NASA CR-134991, National Aeronautics and Space Administration, Lewis Research Center, Cleveland, Ohio 1976.

6. Pinckert, R. E., "Damage Tolerance Assessment of F-4 Aircraft," AIAA-76-904, Presented at AIAA Aircraft Systems and Technology Meeting, Dallas, Texas, September 1976.

7. Rudd, J. L. and T. D. Gray, "Quantification of Fastener-Hole Quality," Journal of Aircraft, Volume 15, Number 13, March 1978.

8. Potter, J. M., "Advances in Fastener Hole Quality Through the Application of Solid Mechanics," Proceedings of the Army Symposium on Solid Mechanics, 1978 - Case Studies on Structural Integrity and Reliability, AMMRC-MS 78-3, Watertown, MA, 1978.

9. Manning, S. D., et al., "Durability Methods Development, Volume 1, Phase I Summary, "Air Force Wright Aeronautical Laboratories, AFFDL-TR-79-3118, Wright-Patterson Air Force Base, Ohio, September 1979.

10. Yang, J. N., "Statistical Estimation of Service Cracks and Maintenance Cost for Aircraft Structures," Journal of Aircraft, Volume 13, Number 12, December 1976.

11. Berens, A. P., "Predicted Crack Repair Costs for Aircraft Structures," Aeronautical Systems Division, ASD-TR-78-39, Wright-Patterson Air Force Base, Ohio, November 1978.

12. Williams, D. A., "Extra Binomial Variation in Logistic Linear Models," Applied Statistics, Royal Statistical Society, Volume 31, Number 2, pp. 144-148, 1982.

13. Cheng, R. C. H. and T. C. Iles, "Confidence Bands for Cumulative Distribution Functions of Continuous Random Variables," Technometrics, Volume 25, No. 1, pp. 77-86, 1983.

14. Anderson, R. L. and T. A. Bancroft, Statistical Theory in Research, McGraw-Hill Book Company, Inc., New York, 1952, pp. 325-330.

15. Wood, Howard, A. and Robert M. Engle, Jr., "USAF Damage Tolerant Design Handbook: Guidelines for the Analysis and Design of Damage Tolerant Aircraft," AFFDL-TR-79-3021, Air Force Wright Aeronautical Laboratories, Wright-Patterson Air Force Base, Ohio, 45433, March 1979.

16. "Final Report on the Effectiveness of the Air Force Nondestructive Inspection Program," Panel on Nondestructive Inspection, Air Force Studies Board, National Research Council, 2101 Constitution Avenue, N. W., Washington, D. C. 20418, 1983.

17. Cox, D. R., The Analysis of Binary Data, Methuen and Co., LTD, London, 1970.

APPENDIX A
STATISTICAL ESTIMATION OF POD

This appendix provides the details of some analysis techniques for estimating POD functions. The first section provides two techniques for analyzing NDI results when recorded only in the pass/fail form. The second section describes the estimation of POD from $\hat{a}$ versus a data.

A.1 ANALYSIS OF PASS/FAIL DATA

There are two techniques that can be used to analyse pass/fail data, depending on the type of data. A regression analysis can be used to estimate the parameters of the POD model when there are multiple inspections of each flaw or a large number of flaws. For data with a single inspection per flaw, the maximum likelihood method provides good estimates of the POD model parameters.

The analyses described in this section are based on the log logistic function given in equation (8). A direct analysis of the model when expressed in the form given by equation (8) is very complicated. The analysis can be simplified by using the re-parameterized model.

$$POD(a) = \frac{\exp(\alpha + \beta\, \ell n(a))}{1 + \exp(\alpha + \beta\, \ell n(a))} \qquad (A1)$$

The relationship between μ and σ of equation (8) and α and β of equation A1 is:

$$\mu = -\alpha/\beta \qquad (A2)$$

$$\sigma = \pi/(\beta\sqrt{3}) \qquad (A3)$$

For both the regression technique and the maximum likelihood method, estimates of μ and σ can be calculated by substituting the appropriate estimates of α and β into the right hand sides of equations (A2) and (A3).

A.1.1 Regression Analysis

A common name for the log logistic model is the log odds model which comes from the logarithm of the odds (p/(1-p)) (log odds) transformation. The log odds transformation converts equation (A1) to

$$\ln \left(\frac{POD(a)}{1-POD(a)}\right) = \alpha + \beta \ln a, \tag{A4}$$

which is linear in the transformed variables

$$Y(a) = \ln \left(\frac{POD(a)}{1-POD(a)}\right) \text{ and } X = \ln a \tag{A5}$$

Linear regression methods are then used to estimate α and β.

Before performing a linear regression on NDI reliability data, the data must be reduced to a set of n pairs, (a_i, p_i), where a_i is the crack length for the ith pair and p_i is the proportion of times the flaw (or flaws) were detected. If the data contain multiple inspections of each flaw, a_i will be the length of a single flaw and p_i will be the proportion of time that the flaw was detected. If flaws are grouped into crack length intervals, a_i will be the midpoint of the ith interval and p_i will be the proportion of flaws in the ith interval that were detected.

Given the n pairs of (a_i, p_i) data points to be fit by the regression analysis, the transformations of equation (A5) are performed, resulting in a set of n (X_i, Y_i) pairs.

The variables X and Y are then used in a linear regression analysis resulting in estimates $\hat{\alpha}$ and $\hat{\beta}$ for α and β, respectively. The formulas for $\hat{\alpha}$ and $\hat{\beta}$ are

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} \tag{A6}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\,\bar{X} \tag{A7}$$

where $\bar{Y}$ and $\bar{X}$ are given by

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}, \quad \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}. \tag{A8}$$

The estimated mean Y as a function of a is given by

$$\bar{Y}(a) = \hat{\alpha} + \hat{\beta}\,\ell n\, a \tag{A9}$$

The formula for a lower confidence bound on the mean μ_{y_L} for a given value is

$$\bar{Y}_L = \hat{\alpha}+\hat{\beta}\,X - t_{(n-2)\gamma}(S)\,\sqrt{\frac{1}{n} + \frac{(X-\bar{X})^2}{SSX}} \tag{A10}$$

where

γ is the confidence coefficient

$t_{(n-2),\gamma}$ is the γth percentile of a t distribution with n-2 degrees of freedom

$$S = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - A - B\,X_i)^2} \tag{A11}$$

$$SSX = \sum_{i=1}^{n} X_i^2 - \frac{\sum_{i=1}^{n} X_i^2}{n} \tag{A12}$$

The inverse Y transformation applied to equation (A9) gives the estimate of the POD and since the log odds transformation is monotonic, the inverse Y transformation of $\bar{Y}_L$ gives the confidence bound on the POD function. The equations for the estimate of POD(a) and its lower confidence bound are:

$$POD(a) = \frac{\exp(\bar{Y}(a))}{1 + \exp(\bar{Y}(a))} \tag{A13}$$

and

$$POD_L(a) = \frac{\exp(\bar{Y}_L(a))}{1 + \exp(\bar{Y}_L(a))} \tag{A14}$$

A problem in the use of regression analysis arises when the observed proportion of detected cracks at a crack length is zero or one. In either of these cases, the most useful transformations can be undefined. To circumvent this problem, there are several alternatives. In the Have Cracks Study [1], the vaues of 0.01 and 0.999 were substituted for 0 and 1, respectively. However, the regression results are sensitive to the arbitrarily defined values. A more acceptable solution is to use a different estimator for the detection probability.

The usual estimator for the detection probability is taken as

$$\hat{p} = i/n \tag{A15}$$

where i is the number of detections and n is the number of specimens with the crack of the fixed length. Other estimates of the proportion which have acceptable statistical properties are the mean estimate

$$\bar{p} = \begin{cases} \frac{i}{n+i} & \text{if } i > \frac{n}{2} \\ \frac{1}{2} & \text{if } i = n/2 \\ \frac{i+1}{n+i} & \text{if } i < \frac{n}{2} \end{cases} \tag{A16}$$

and the median estimate:

$$\tilde{p} = \begin{cases} \frac{i-0.3}{n+0.4} & \text{if } i > \frac{n}{2} \\ \frac{1}{2} & \text{if } i = n/2 \\ \frac{i+0.7}{n+0.4} & \text{if } i < \frac{n}{2} \end{cases} \qquad \text{(A17)}$$

The use of either equation A16 or equation A17 for observed detection probabilities would eliminate an arbitrary definition when the observed proportion of detections is zero or one.

A.1.2 Maximum Likelihood Estimates

Given the POD(a) model of equation (A1), an entirely different method for estimating the parameters uses the principal of maximum likelihood. In this type of estimation the parameter estimates are the values which maximize the probability of obtaining the observed data. The maximum likelihood estimates do not require grouping of data when the experiment involves a single inspection per crack. Instead, they are based directly on the observed outcomes of 0 for a non-detection and 1 for a detection. This paragraph presents the equations for the maximum likelihood estimates of the log odds model and confidence limits when each crack is inspected only once. Maximum likelihood estimates for multiple inspections of each crack could also be developed. Further, maximum likelihood estimators for parameters of models other than equation (A1) could be developed, but the solutions would not necessarily be in closed form.

Maximum likelihood estimation is based on the concept that the data will take on values which are most likely to occur under the chosen probability model. For example, in a simple Bernoulli trial (which is the probabilistic representation of a single inspection) the probability of success is p. If $p > 1/2$ a success would be more likely than a failure. Conversely, if a success were observed in one trial, it is more likely that

$p > 1/2$. In the following, the philosophy of maximum likelihood estimation, the value of the unknown parameter that would give rise to the highest probability of obtaining the observed data is used as the estimate. In the simple Bernoulli experiment if p were equal to 1 the probability of observing a success would be 1. Since probability cannot exceed 1, the maximum likelihood estimate of p when a success is observed in a single Bernoulli trial is 1.

To find the maximum likelihood estimates of equation (A1) from a sample of single inspections of n cracks, the following procedure adopted from Cox[17] can be used. The maximum likelihood estimates $\hat{\alpha}$ and $\hat{\beta}$ of α and β satisfy the simultaneous equations.

$$0 = \sum_{i=1}^{n} z_i - \sum_{i=1}^{n} \frac{\exp(\hat{\alpha}+\hat{\beta}\ell n(a_i))}{1+\exp(\hat{\alpha}+\hat{\beta}\ell n(a_i))} \tag{A18}$$

$$0 = \sum_{i=1}^{n} z_i \ell n(a_i) - \sum_{i=1}^{n} \frac{\ell n(a_i)\exp(\hat{\alpha}+\hat{\beta}\ell n(a_i))}{1+\exp(\hat{\alpha}+\hat{\beta}\ell n(a_i))} \tag{A19}$$

where $z_i = 1$ if the flaw is detected and 0 if it is not. The variances and covariance of the estimates $\hat{\alpha}$ and $\hat{\beta}$ are

$$\text{Var}\ (\hat{\alpha}) = \sum_{i=1}^{n} \frac{\exp(\alpha+\beta\ell n(a_i))}{(1+\exp(\alpha+\beta\ell n(a_i)))} \tag{A20}$$

$$\text{Var}\ (\hat{\beta}) = \sum_{i=1}^{n} \frac{(\ell n(a_i))^2\exp(\alpha+\beta\ell n(a_i))}{(1+\exp(\alpha+\beta\ell n(a_i)))^2} \tag{A21}$$

$$\text{Cov}\ (\hat{\alpha},\hat{\beta}) = \sum_{i=1}^{n} \frac{\ell n(a_i)\exp(\alpha+\beta\ell n(a_i))}{(1+\exp(\alpha+\beta\ell n(a_i)))^2} \tag{A22}$$

Estimates of these variances and covariance are calculated by substituting the estimates, $\hat{\alpha}$ and $\hat{\beta}$, in equations (A20), (A21), and (A22).

The maximum likelihood estimate of the POD function is calculated by substituting $\hat{\alpha}$ and $\hat{\beta}$ for α and β in equation (A1). The change of variables must be made using the same transformation that was used in the regression analysis to obtain

$$\log\,(p(a)/(1-p(a))) = Y(a) = \hat{\alpha} + \hat{\beta}\ \ell n\ (a_i) \tag{A23}$$

For very large sample sizes, estimates of the variances and covariance of $\hat{\alpha}$ and $\hat{\beta}$ can be used to calculate a lower confidence bound on Y(a) as given by

$$Y_L(a) = \hat{\alpha} + \hat{\beta}\ \ell n(a) - Z_\gamma \sqrt{S^2_\alpha + 2\ell n(a)\ S^2_{\hat{\alpha}\hat{\beta}} + (\ell n(a))^2 S^2_{\hat{\beta}}} \tag{A24}$$

where

γ is the confidence level,

Z_γ satisfies $P(Z < Z_\gamma) = \gamma$ for the standard normal distribution

$S^2_{\hat{\alpha}}$ is the estimate of Var $(\hat{\alpha})$,

$S^2_{\hat{\alpha}\hat{\beta}}$ is the estimate of Cov $(\hat{\alpha},\hat{\beta})$,

$S^2_{\hat{\beta}}$ is the estimate of Var $(\hat{\beta})$.

Since the log odds transformation is monotonic, the reverse transformation of the confidence bound on Y(a) will be the confidence bound on P(a). Specifically,

$$POD(a) = \frac{\exp(Y(a))}{1+\exp(Y(a))} \tag{A25}$$

$$POD_L(a) = \frac{\exp(Y_L(a))}{1+\exp(Y_L(a))} \tag{A26}$$

Generally, maximum likelihood estimates are better than regression estimates from grouped data; however, if the number of flaws is very large (greater than 100) and the groupings do not result in many 0's and 1's for p_i's, the results of the analyses should be similar.

In inadequately designed NDI reliability experiments it is possible to obtain estimates of $\hat{\beta}$ which are not significantly greater than zero. See Subsection 3.3.2.1 for a discussion of this problem.

A2. METHOD OF MOMENTS ESTIMATORS

Suitable initial estimates of the POD model parameters are necessary in applying the maximum likelihood method. When the maximum likelihood equations have more than one solution, the computer algorithm converges to the solution closest to the initial estimates. For estimating POD parameters, the initial estimators were chosen to be those as obtained from the method of moments. The following paragraphs describe the application of the method of moments to the log odds model.

The method of moments is an algebraically simple computational technique that is first used to calculate initial estimates of μ and σ. Equations A2 and A3 convert the estimates of μ and σ to α and β for use in the computer algorithm. The method of moments estimates the parameters of a distribution function by equating the sample moments calculated from data to the theoretical moments expressed in terms of the parameters of the distribution. The log logistic function that is being used to model the POD function is actually a cumulative distribution function and the μ and σ (of the functional parametrization of equation 8) are the mean and standard deviation, respectively.

The application of the method of moments to NDI reliability data is not straightforward. Direct measurements of random variables with POD(a) as their distribution function are not made. The only information available is the set of n (a_i, p_i) pairs. However, this information can be used in the context of a non-standard representation of the mean and variance of a distribution as follows.

The standard definition for the mean of a distribution, say F(x) is

$$\mu = \int_{-\infty}^{\infty} x \, dF(x) \qquad \text{(A29)}$$

Integration by parts can convert equation (A29) to

$$\mu = \int_{0}^{\infty} [1-F(x)] \, dx - \int_{-\infty}^{0} F(x) \, dx \qquad \text{(A30)}$$

In the log logistic function, μ is the expected value of ln(a) so that the X transformation of equation (A5) must be used before applying equation (A30). The log logistic function of equation (8) is therefore converted to

$$POD(x) = [1 + \exp (\frac{-\pi}{\sqrt{3}} (\frac{x-\mu}{\sigma}))] \, . \qquad \text{(A31)}$$

In terms of the POD function, equation (A30) is illustrated in Figure A1. In Figure A1, area A represents the first integral of equation (A30) and area B represents the second integral. The mean, μ, is area A minus area B. To apply the method of moments to NDI reliability data, the analog of Figure A1 is drawn for a set by NDI reliability data as Figure A2.

Figure A2 is a plot of the (x_i, p_i) pairs, where x_i is the log transform of a_i. The expected value of p_i is $POD(x_i)$ so that the curve consisting of the line segments connecting the (x,p) pairs ordered by increasing x is an approximation of the function POD(x). Areas C and D of Figure A2 correspond to areas A and B of Figure A1 respectively so that the method of moments estimate of μ is

$$\tilde{\mu} = C - D \qquad \text{(A32)}$$

The calculation of the method of moments estimate of μ can be simplified by choosing a scale so that all crack lengths are greater than 1 and therefore all x values are greater than 0. Crack lengths given in the units of mils usually satisfy this

criteria. When all X values are greater than 0, the area D will be 0 so that only area C need be calculated. The formula for the method of moments estimate of μ then becomes

$$\tilde{\mu} = X_n - x_1 - \frac{1}{2} \sum_{i=2}^{n} (x_i - x_{i-1})(p_i + p_{i-1}) \qquad \text{(A33)}$$

where x's have been ordered so that $X_1 < X_2 < \ldots < X_n$. The ordering should be strict so that if several flaws have the same length, they should be grouped and p taken as the total proportion of all inspections on all flaws of that length that were detections.

A similar process can be used to show that the method of moments estimate of σ is

$$\tilde{\sigma} = \sqrt{\sqrt{X_n} - \sqrt{X_1} - \frac{1}{2} \sum_{i=2}^{n} \{ (\sqrt{X_i} - \sqrt{X_{i-1}})(p_i + p_{i-1}) \} - \tilde{\mu}^2} \qquad \text{(A34)}$$

The statistical properties of the method of moments estimators of μ and σ depend on how well a numerical integration of the true POD function based on the crack lengths of the sample data approximates the true values of μ and σ. If the true POD can be reasonably approximated prior to the NDI reliability experiment, the experiment could be designed so that the method of moments provided good estimates of the POD parameters. However, in most applications the method of moments should be used only to provide initial estimates for the maximum likelihood algorithm.

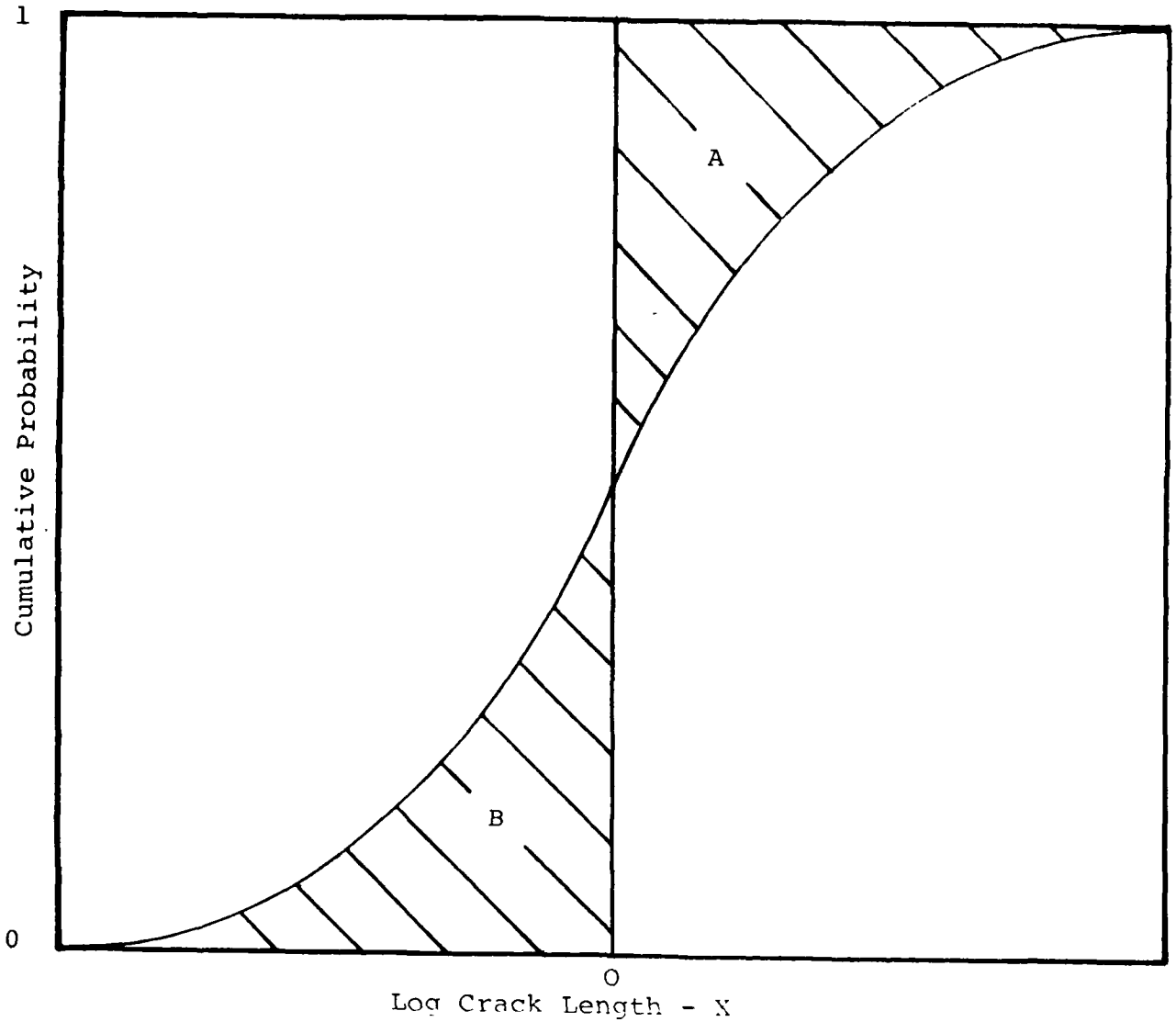


Figure A1. Illustration of Alternate Formula for the Mean of a Distribution.

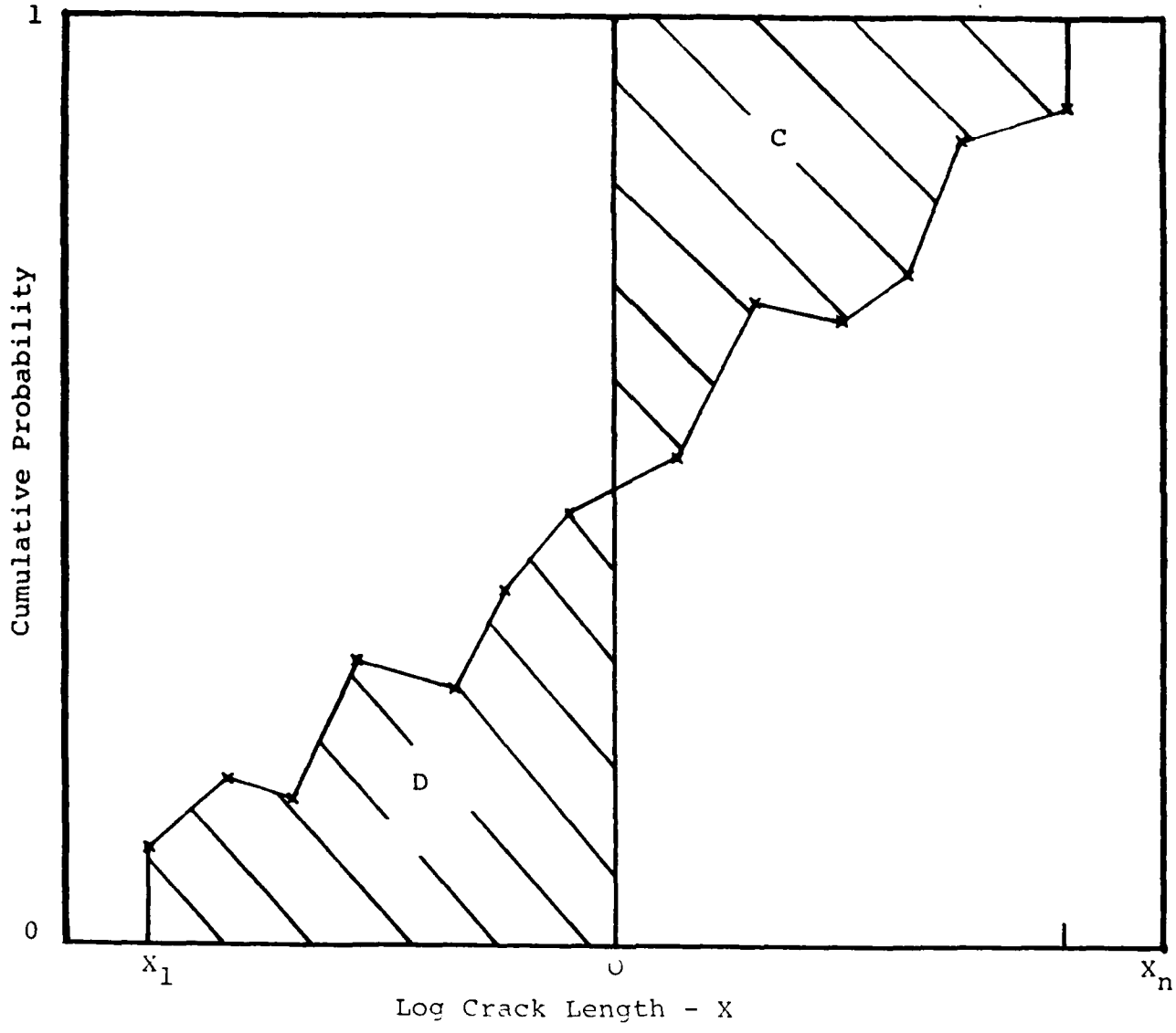
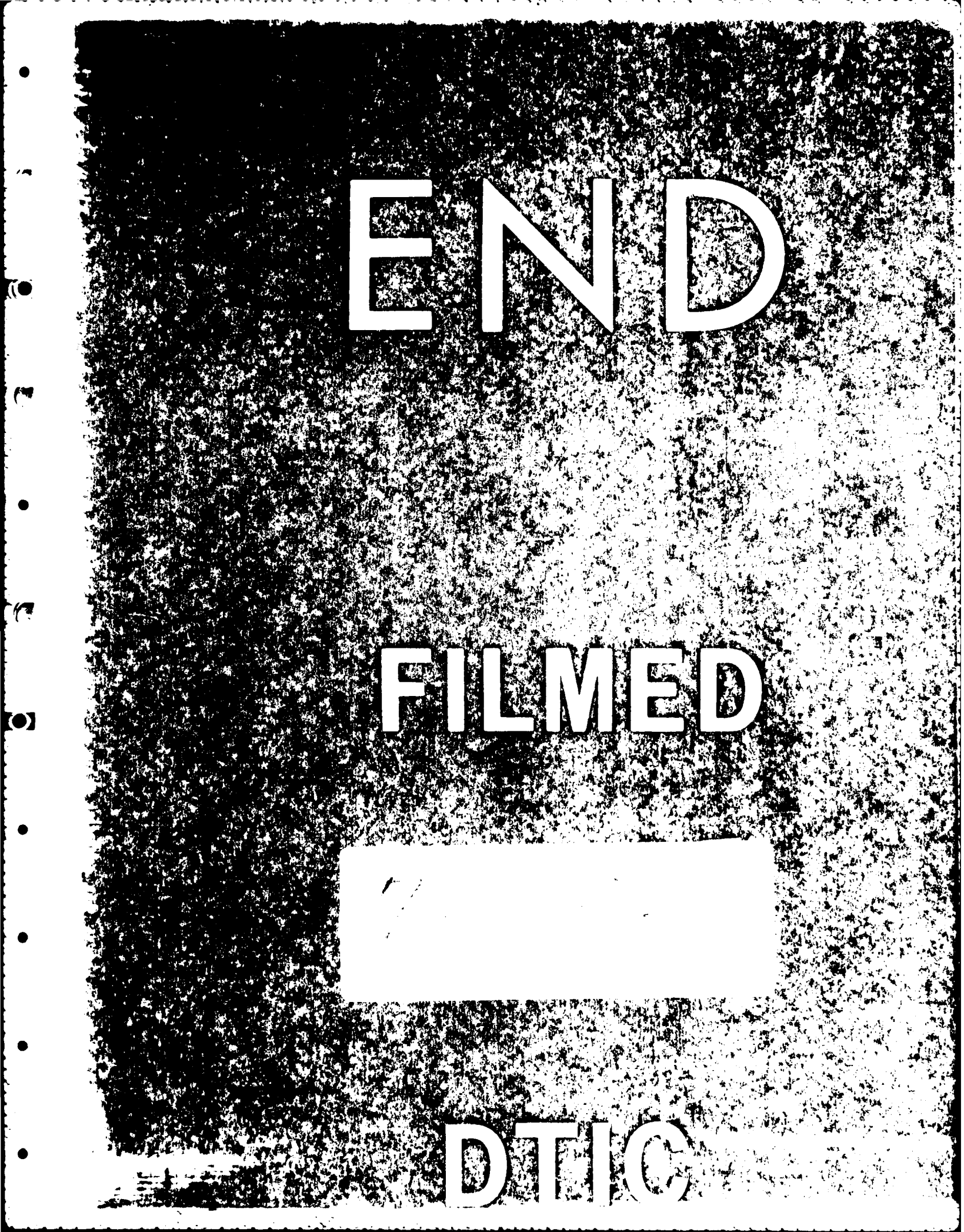


Figure A2. Analogous Areas for Observed POD Function.


END
FILMED
DTIC